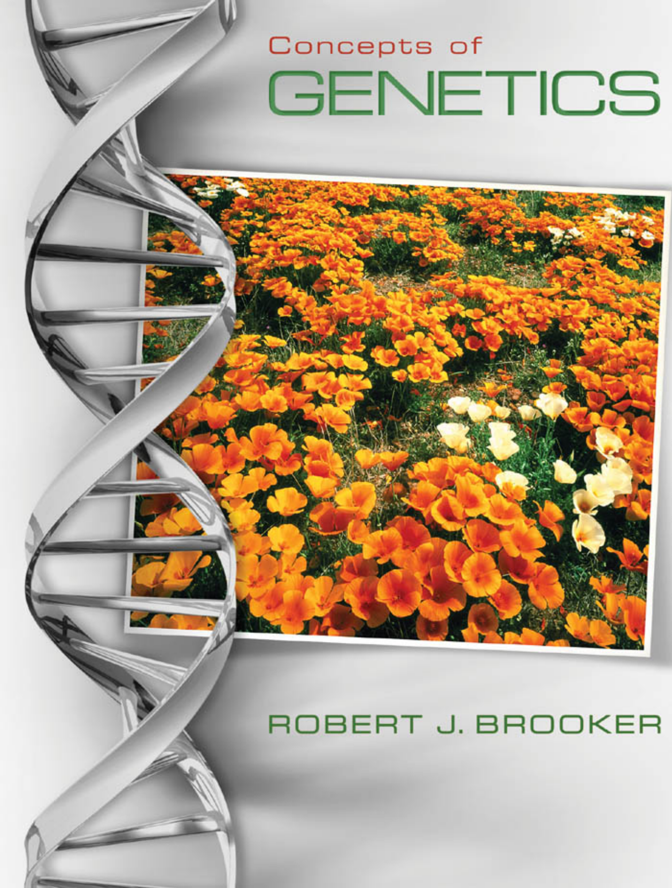


Concepts of
GENETICS

ROBERT J. BROOKER

VISIT…

LANZAROTE
Caliente.COM

This page intentionally left blank

Concepts of
GENETICS

ROBERT J. BROOKER

McGraw Hill
Connect
Learn
Succeed™

CONCEPTS OF GENETICS

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2012 by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOW/DOW 1 0 9 8 7 6 5 4 3 2 1

ISBN 978–0–07–352533–4
MHID 0–07–352533–2

Vice President, Editor-in-Chief: *Marty Lange*
Vice President, EDP: *Kimberly Meriwether David*
Senior Director of Development: *Kristine Tibbetts*
Publisher: *Janice Roerig-Blong*
Director of Development: *Elizabeth M. Sievers*
Developmental Editor: *Mandy C. Clark*
Executive Marketing Manager: *Patrick E. Reidy*
Senior Project Manager: *Jayne L. Klein*
Buyer II: *Sherry L. Kane*
Senior Media Project Manager: *Tammy Juran*
Senior Designer: *David W. Hash*
Cover Designer: *John Joran*
Cover Image: *Mexican Gold Poppy (Eschscholzia mexicana) field near Pinaleno Mountains, Cochise County, Arizona ©Willard Clay/Oxford Scientific/Getty Images; Model of DNA ©Thinkstock/Herma Collection*
Senior Photo Research Coordinator: *John C. Leland*
Photo Research: *Copyright Works Inc.*
Compositor: *Lachina Publishing Services*
Typeface: *10/12 Minion*
Printer: *R. R. Donnelley*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Library of Congress Cataloging-in-Publication Data

Brooker, Robert J.
Concepts of genetics / Robert J. Brooker. -- 1st ed.
p. cm.
Includes index.
ISBN 978–0–07–352533–4 — ISBN 0–07–352533–2 (hard copy : alk. paper) 1. Genetics. I. Title.
QH430.B764 2012
576.5--dc22
2010042316

www.mhhe.com

BRIEF CONTENTS

PART I

1 Overview of Genetics 1

PART II

2 Reproduction and Chromosome Transmission 19
3 Mendelian Inheritance 40
4 Sex Determination and Sex Chromosomes 71
5 Extensions of Mendelian Inheritance 88
6 Extranuclear Inheritance, Imprinting, and Maternal Effect 110
7 Genetic Linkage and Mapping in Eukaryotes 129
8 Variation in Chromosome Structure and Number 155
9 Genetics of Bacteria 185
10 Genetics of Viruses 206

PART III

11 Molecular Structure of DNA and RNA 225
12 Molecular Structure and Organization of Chromosomes 246
13 DNA Replication and Recombination 268

PART IV

14 Gene Transcription and RNA Modification 299
15 Translation of mRNA 324
16 Gene Regulation in Bacteria 355
17 Gene Regulation in Eukaryotes 379
18 Gene Mutation and DNA Repair 411

PART V

19 Recombinant DNA Technology 443
20 Biotechnology 474
21 Genomics I: Analysis of DNA and Transposable Elements 499
22 Genomics II: Functional Genomics, Proteomics, and Bioinformatics 531

PART VI

23 Medical Genetics and Cancer 551
24 Developmental Genetics and Immunogenetics 585
25 Population Genetics 614
26 Quantitative Genetics 646
27 Evolutionary Genetics 672

TABLE OF CONTENTS

Preface viii

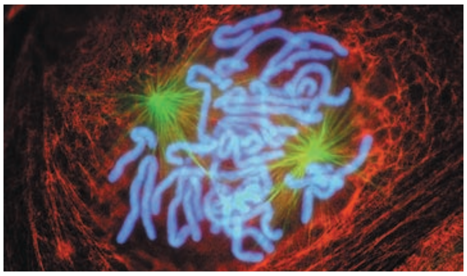
A Visual Guide to *Concepts of Genetics* xiv

PART I

INTRODUCTION 1

1 OVERVIEW OF GENETICS 1

1.1 The Molecular Expression of Genes 4
1.2 The Relationship between Genes and Traits 7
1.3 Fields of Genetics 12
Key Terms 16
Chapter Summary 16
Problem Sets & Insights 16



PART II

PATTERNS OF INHERITANCE 19

2 REPRODUCTION AND CHROMOSOME TRANSMISSION 19

2.1 General Features of Chromosomes 19
2.2 Cell Division 23
2.3 Mitosis and Cytokinesis 26
2.4 Meiosis 30
2.5 Sexual Reproduction 34
Key Terms 37
Chapter Summary 37
Problem Sets & Insights 37

3 MENDELIAN INHERITANCE 40

3.1 Mendel's Study of Pea Plants 41
3.2 Law of Segregation 44

Experiment
Mendel Followed the Outcome of a Single Character for Two Generations 44

3.3 Law of Independent Assortment 48

Experiment
Mendel Also Analyzed Crosses Involving Two Different Characters 48

3.4 Chromosome Theory of Inheritance 53
3.5 Studying Inheritance Patterns in Humans 56
3.6 Probability and Statistics 57
Key Terms 63
Chapter Summary 63
Problem Sets & Insights 64

4 SEX DETERMINATION AND SEX CHROMOSOMES 71

4.1 Mechanisms of Sex Determination Among Various Species 71
4.2 Dosage Compensation and X Inactivation in Mammals 75
4.3 Properties of the X and Y Chromosome in Mammals 78
4.4 Transmission Patterns for X-Linked Genes 79

Experiment
Morgan's Experiments Showed a Connection Between a Genetic Trait and the Inheritance of a Sex Chromosome in *Drosophila* 79

Key Terms 83
Chapter Summary 83
Problem Sets & Insights 84

5 EXTENSIONS OF MENDELIAN INHERITANCE 88

5.1 Overview of Simple Inheritance Patterns 88
5.2 Dominant and Recessive Alleles 90
5.3 Environmental Effects on Gene Expression 92
5.4 Incomplete Dominance, Overdominance, and Codominance 94
5.5 Sex-Influenced and Sex-Limited Inheritance 98
5.6 Lethal Alleles 100
5.7 Pleiotropy 101
5.8 Gene Interactions 102
Key Terms 105
Chapter Summary 105
Problem Sets & Insights 106

6 EXTRANUCLEAR INHERITANCE, IMPRINTING, AND MATERNAL EFFECT 110

6.1 Extranuclear Inheritance: Chloroplasts 110
6.2 Extranuclear Inheritance: Mitochondria 114
6.3 Theory of Endosymbiosis 116
6.4 Epigenetics: Imprinting 118
6.5 Maternal Effect 123
Key Terms 125
Chapter Summary 125
Problem Sets & Insights 126

7 GENETIC LINKAGE AND MAPPING IN EUKARYOTES 129

7.1 Overview of Linkage 129
7.2 Relationship Between Linkage and Crossing Over 131
7.3 Genetic Mapping in Plants and Animals 137

Experiment
Alfred Sturtevant Used the Frequency of Crossing Over in Dihybrid Crosses to Produce the First Genetic Map 140

7.4 Mitotic Recombination 145
Key Terms **147**
Chapter Summary **147**
Problem Sets & Insights **148**

8 VARIATION IN CHROMOSOME STRUCTURE AND NUMBER 155

8.1 Microscopic Examination of Eukaryotic Chromosomes 155
8.2 Changes in Chromosome Structure: An Overview 158
8.3 Deletions and Duplications 159
8.4 Inversions and Translocations 162
8.5 Changes in Chromosome Number: An Overview 168
8.6 Variation in the Number of Chromosomes Within a Set: Aneuploidy 169
8.7 Variation in the Number of Sets of Chromosomes 171
8.8 Mechanisms That Produce Variation in Chromosome Number 174
Key Terms **180**
Chapter Summary **180**
Problem Sets & Insights **181**

9 GENETICS OF BACTERIA 185

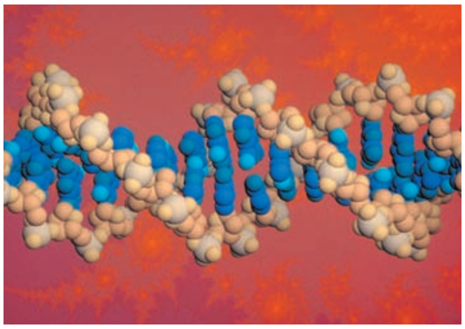
9.1 Overview of Genetic Transfer in Bacteria 186
9.2 Bacterial Conjugation 187
9.3 Conjugation and Mapping via *Hfr* Strains 191

Experiment
Conjugation Experiments Can Map Genes Along the *E. coli* Chromosome 193

9.4 Bacterial Transduction 196
9.5 Bacterial Transformation 199
Key Terms **201**
Chapter Summary **201**
Problem Sets & Insights **201**

10 GENETICS OF VIRUSES 206

10.1 Virus Structure and Genetic Composition 207
10.2 Viral Reproductive Cycles 209
10.3 Plaque Formation and Intergenic Complementation in Bacteriophages 215
10.4 Intragenic Mapping in Bacteriophages 218
Key Terms **222**
Chapter Summary **222**
Problem Sets & Insights **223**



PART III
MOLECULAR STRUCTURE AND REPLICATION OF THE GENETIC MATERIAL 225

11 MOLECULAR STRUCTURE OF DNA AND RNA 225

11.1 Identification of DNA as the Genetic Material 225

Experiment
Hershey and Chase Provided Evidence That DNA Is the Genetic Material of T2 Phage 228

11.2 Overview of DNA and RNA Structure 231
11.3 Nucleotide Structure 232
11.4 Structure of a DNA Strand 233
11.5 Discovery of the Double Helix 234
11.6 Structure of the DNA Double Helix 237
11.7 RNA Structure 240
Key Terms **242**
Chapter Summary **242**
Problem Sets & Insights **243**

12 MOLECULAR STRUCTURE AND ORGANIZATION OF CHROMOSOMES 246

12.1 Organization of Sites Along Bacterial Chromosomes 246
12.2 Structure of Bacterial Chromosomes 247
12.3 Organization of Sites Along Eukaryotic Chromosomes 251
12.4 Sizes of Eukaryotic Genomes and Repetitive Sequences 252
12.5 Structure of Eukaryotic Chromosomes in Nondividing Cells 253

Experiment
The Repeating Nucleosome Structure Is Revealed by Digestion of the Linker Region 255

12.6 Structure of Eukaryotic Chromosomes During Cell Division 260
Key Terms **264**
Chapter Summary **264**
Problem Sets & Insights **265**

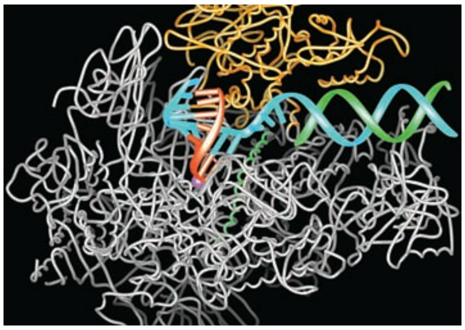
13 DNA REPLICATION AND RECOMBINATION 268

13.1 Structural Overview of DNA Replication 268

Experiment
Three Different Models Were Proposed That Described the Net Result of DNA Replication 270

13.2 Bacterial DNA Replication: The Formation of Two Replication Forks at the Origin of Replication 272
13.3 Bacterial DNA Replication: Synthesis of New DNA Strands 275

13.4 Bacterial DNA Replication: Chemistry and Accuracy 281
13.5 Eukaryotic DNA Replication 283
13.6 Homologous Recombination 287
Key Terms 293
Chapter Summary 293
Problem Sets & Insights 294



PART IV
MOLECULAR PROPERTIES OF GENES 299

14 GENE TRANSCRIPTION AND RNA MODIFICATION 299

14.1 Overview of Transcription 300
14.2 Transcription in Bacteria 302
14.3 Transcription in Eukaryotes 307
14.4 RNA Modification 312
Key Terms 319
Chapter Summary 319
Problem Sets & Insights 320

15 TRANSLATION OF mRNA 324

15.1 The Genetic Basis for Protein Synthesis 324
15.2 The Relationship Between the Genetic Code and Protein Synthesis 327
15.3 Experimental Determination of the Genetic Code 333
Experiment
Synthetic RNA Helped to Determine the Genetic Code 334
15.4 Structure and Function of tRNA 338
15.5 Ribosome Structure and Assembly 341
15.6 Stages of Translation 344
Key Terms 350
Chapter Summary 350
Problem Sets & Insights 351

16 GENE REGULATION IN BACTERIA 355

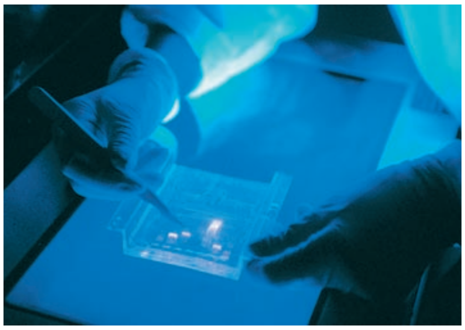
16.1 Overview of Transcriptional Regulation 356
16.2 Regulation of the *lac* Operon 358
Experiment
The *lacI* Gene Encodes a Diffusible Repressor Protein 360
16.3 Regulation of the *trp* Operon 367
16.4 Translational and Posttranslational Regulation 371
16.5 Riboswitches 373
Key Terms 375
Chapter Summary 375
Problem Sets & Insights 375

17 GENE REGULATION IN EUKARYOTES 379

17.1 Regulatory Transcription Factors 380
17.2 Chromatin Remodeling, Histone Variation, and Histone Modification 386
17.3 DNA Methylation 391
17.4 Insulators 394
17.5 Regulation of RNA Processing, RNA Stability, and Translation 396
Experiment
Fire and Mello Show That Double-Stranded RNA Is More Potent Than Antisense RNA at Silencing mRNA 400
Key Terms 405
Chapter Summary 405
Problem Sets & Insights 406

18 GENE MUTATION AND DNA REPAIR 411

18.1 Effects of Mutations on Gene Structure and Function 412
18.2 Random Nature of Mutations 418
18.3 Spontaneous Mutations 421
18.4 Induced Mutations 427
18.5 DNA Repair 430
Key Terms 438
Chapter Summary 438
Problem Sets & Insights 439



PART V
GENETIC TECHNOLOGIES 443

19 RECOMBINANT DNA TECHNOLOGY 443

19.1 Gene Cloning Using Vectors 444
19.2 Polymerase Chain Reaction 450
19.3 DNA Libraries and Blotting Methods 455
19.4 Methods for Analyzing DNA- and RNA-Binding Proteins 462
19.5 DNA Sequencing and Site-Directed Mutagenesis 464
Key Terms 467
Chapter Summary 468
Problem Sets & Insights 468

20 BIOTECHNOLOGY 474

20.1 Uses of Microorganisms in Biotechnology 474
20.2 Genetically Modified Animals 477
20.3 Reproductive Cloning and Stem Cells 483
20.4 Genetically Modified Plants 488
20.5 Human Gene Therapy 491
Experiment
Adenosine Deaminase Deficiency Was the First Inherited Disease Treated with Gene Therapy 492
Key Terms 495
Chapter Summary 496
Problem Sets & Insights 496

21 GENOMICS I: ANALYSIS OF DNA AND TRANSPOSABLE ELEMENTS 499

21.1 Overview of Chromosome Mapping 500
21.2 Cytogenetic Mapping via Microscopy 501
21.3 Linkage Mapping via Crosses 503
21.4 Physical Mapping via Cloning 507
21.5 Genome-Sequencing Projects 512
21.6 Transposition 517
Key Terms 525
Chapter Summary 526
Problem Sets & Insights 526

22 GENOMICS II: FUNCTIONAL GENOMICS, PROTEOMICS, AND BIOINFORMATICS 531

22.1 Functional Genomics 532
22.2 Proteomics 535
22.3 Bioinformatics 540
Key Terms 547
Chapter Summary 547
Problem Sets & Insights 548

PART VI

GENETIC ANALYSIS OF INDIVIDUALS AND POPULATIONS 551

23 MEDICAL GENETICS AND CANCER 551

23.1 Inheritance Patterns of Genetic Diseases 552
23.2 Detection of Disease-Causing Alleles 558
23.3 Prions 563
23.4 Genetic Basis of Cancer 565
Key Terms 578
Chapter Summary 578
Problem Sets & Insights 579

24 DEVELOPMENTAL GENETICS AND IMMUNOGENETICS 585

24.1 Overview of Animal Development 586
24.2 Invertebrate Development 589
24.3 Vertebrate Development 599
24.4 Plant Development 602
24.5 Immunogenetics 605
Key Terms 608
Chapter Summary 608
Problem Sets & Insights 609

25 POPULATION GENETICS 614

25.1 Genes in Populations and the Hardy-Weinberg Equation 614
25.2 Overview of Microevolution 619
25.3 Natural Selection 620

Experiment
The Grants Have Observed Natural Selection in Galápagos Finches 626

25.4 Genetic Drift 628
25.5 Migration 630
25.6 Nonrandom Mating 631
25.7 Sources of New Genetic Variation 633
Key Terms 638
Chapter Summary 638
Problem Sets & Insights 639

26 QUANTITATIVE GENETICS 646

26.1 Overview of Quantitative Traits 646
26.2 Statistical Methods for Evaluating Quantitative Traits 648
26.3 Polygenic Inheritance 651
26.4 Heritability 656

Experiment
Heritability of Dermal Ridge Count in Human Fingerprints Is Very High 659

26.5 Selective Breeding 663
Key Terms 666
Chapter Summary 666
Problem Sets & Insights 667

27 EVOLUTIONARY GENETICS 672

27.1 Origin of Species 673
27.2 Phylogenetic Trees 679
27.3 Molecular Evolution and Molecular Clocks 686
27.4 Evo-Devo: Evolutionary Developmental Biology 692
Key Terms 696
Chapter Summary 696
Problem Sets & Insights 697

Appendix A
Experimental Techniques A-1

Appendix B
Solutions to Concept Checks and Even-Numbered Problems A-7

Glossary G-1

Credits C-1

Index I-1

PREFACE

Based on our discussions with instructors from many institutions, I have learned that most instructors want a broad textbook that clearly explains concepts in a way that is interesting, accurate, concise, and up-to-date. *Concepts of Genetics* has been written to achieve these goals. It is intended for students who want to gain a conceptual grasp of the various fields of genetics. The content reflects current trends in genetics and the pedagogy is based on educational research. In particular, a large amount of formative assessment is woven into the content. As an author, researcher, and teacher, I want a textbook that gets students actively involved in learning genetics. To achieve this goal, I have worked with a talented team of editors, illustrators, and media specialists who have helped me to make the first edition of *Concepts of Genetics* a fun learning tool. The features that we feel are most appealing to students are the following.

- **Formative assessment** Perhaps the most difficult challenge for each student is to figure out what it is they don't know or don't fully understand. Formative assessment is often a self-reflective process in which a student answers questions and the feedback from those questions allows her or him to recognize the status of their learning. When it works well, it helps to guide a student through the learning process. In *Concepts of Genetics*, a student is given formative assessment in multiple ways. First, most of the figure legends contain "Concept check" questions that test a student's understanding of the material. The answers to these questions are provided in the back of the book, so the student can immediately determine if their own answer is correct. Second, the end of each section of each chapter contains multiple choice questions that test the broader concepts that were described in that section. The answers are at the end of the chapter, which allows for immediate feedback for the student. Third, a rigorous set of problems is provided at the end of each chapter. These problem sets are divided into Conceptual questions, Application and Experimental questions, and Questions for Student Discussion/Collaboration.
- **Chapter organization** In genetics, it is sometimes easy to "lose the forest for the trees." Genetics is often times a dense subject. To circumvent this difficulty, the content in *Concepts of Genetics* has been organized to foster a better appreciation for the big picture of genetic principles. The chapters are divided into several sections, and each section ends with a summary that touches on the main points. As mentioned, multiple choice questions at the end of each section are also intended to help students grasp the broader concepts in genetics. Finally, the end of each chapter contains a summary, which allows students to connect the concepts that were learned in each section.
- **Connecting molecular genetics and traits** It is commonly mentioned that students often have trouble connecting the concepts they have learned in molecular genetics with the traits that occur at the level of a whole organism (i.e., What does transcription have to do with blue eyes?). To try to make this connection more meaningful, certain figure legends in each chapter, designated **Genes→Traits**, remind students that molecular and cellular phenomena ultimately lead to the traits that are observed in each species.
- **Interactive exercises** Working with education specialists, the author has crafted interactive exercises in which the students can make their own choices in problem-solving activities and predict what the outcomes will be. Many of these exercises are focused on inheritance patterns and human genetic diseases. (For example, see Chapters 5 and 23.) In addition, we have many interactive exercises for the molecular chapters. These types of exercises engage students in the learning process. The interactive exercises are found online and the corresponding material in the chapter is indicated with an Interactive Exercise icon.

  INTERACTIVE EXERCISE
- **Animations** Our media specialists have created over 50 animations for a variety of genetic processes. These animations were made specifically for this textbook and use the art from the textbook. The animations literally make many of the figures in the textbook "come to life." The animations are found online and the corresponding material in the chapter is indicated with an Online Animation icon.

  ONLINE ANIMATION
- **Experiments** Many chapters have an experiment that is presented according to the scientific method. These experiments are not "boxed off" from the rest of the chapter. Rather, they are integrated within the chapters and flow with the rest of the text. As you are reading the experiments, you will simultaneously explore the scientific method and the genetic principles that have been discovered using this approach. For students, I hope this textbook helps you to see the fundamental connection between scientific analysis and principles. For both students and instructors, I expect that this strategy makes genetics much more fun to explore.
- **Art** A large proportion of a student's efforts is aimed at studying figures. As described later in this preface, the art is clearly a strength of this textbook. Most of the work in producing this book has gone into the development of the art. It is designed to be complete, clear, consistent, and realistic.

- **Engaging text** A strong effort has been made to pepper the text with questions. Sometimes these are questions that scientists considered when they were conducting their research. Sometimes they are questions that the students might ask themselves when they are learning about genetics.

Overall, an effective textbook needs to accomplish three goals. First, it needs to provide comprehensive, accurate, and up-to-date content in its field. Second, it needs to expose students to the techniques and skills they will need to become successful in that field. And finally, it should inspire students so they want to pursue that field as a career. The hard work that has gone into the first edition of *Concepts of Genetics* has been aimed at achieving all three of these goals.

HOW WE EVALUATED YOUR NEEDS

ORGANIZATION

In surveying many genetics instructors, it became apparent that most people fall into two camps: **Mendel first** versus **Molecular first**. I have taught genetics both ways. As a teaching tool, this textbook has been written with these different teaching strategies in mind. The organization and content lend themselves to various teaching formats.

Chapters 2 through 10 are largely inheritance chapters, whereas Chapters 25 through 27 examine population and quantitative genetics. The bulk of the molecular genetics is found in Chapters 11 through 24, although I have tried to weave a fair amount of molecular genetics into Chapters 2 through 10 as well. The information in Chapters 11 through 24 does not assume that a student has already covered Chapters 2 through 10. Actually, each chapter is written with the perspective that instructors may want to vary the order of their chapters to fit their students' needs.

For those who like to discuss inheritance patterns first, a common strategy would be to cover Chapters 1 through 10 first, and then possibly 25 through 27. (However, many instructors like to cover quantitative and population genetics at the end. Either way works fine.) The more molecular and technical aspects of genetics would then be covered in Chapters 11 through 24. Alternatively, if you like the "Molecular first" approach, you would probably cover Chapter 1, then skip to Chapters 11 through 24, then return to Chapters 2 through 10, and then cover Chapters 25 through 27 at the end of the course. This textbook was written in such a way that either strategy works well.

ACCURACY

Both the publisher and I acknowledge that inaccuracies can be a source of frustration for both the instructor and students. Therefore, throughout the writing and production of this textbook we have worked very hard to catch and correct errors during each phase of development and production.

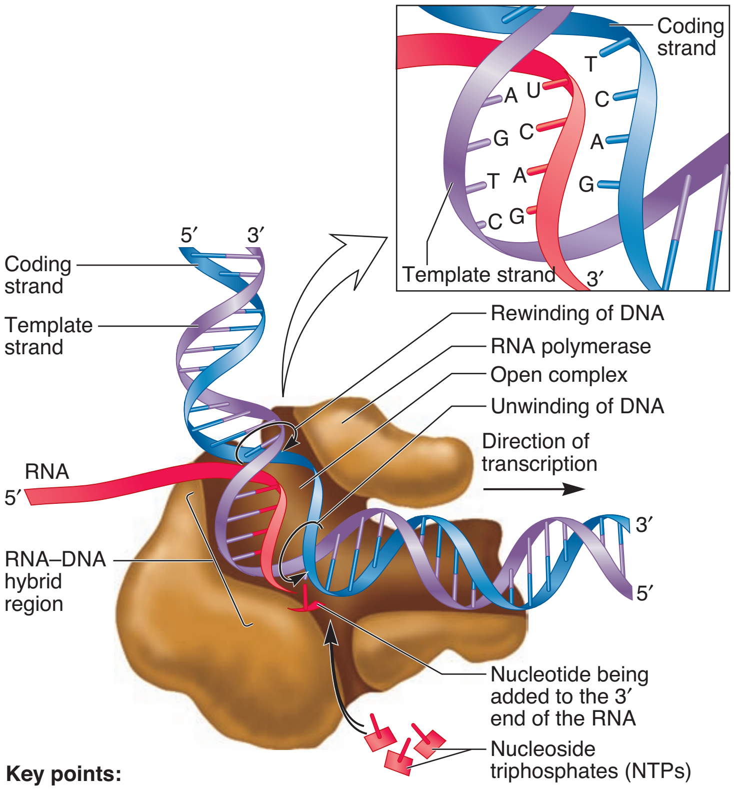
Each chapter has been reviewed by a minimum of 8 people. At least 6 of these people are faculty members who teach the course or conduct research in genetics or both. In addition, a developmental editor has gone through the material to check for accuracy in art and consistency between the text and art. When they were first developed, we had a team of students work through all of the problem sets and one development editor also checked them. The author personally checked every question and answer when the chapters were completed.

ILLUSTRATIONS

In surveying students whom I teach, I often hear it said that most of their learning comes from studying the figures. Likewise, instructors frequently use the illustrations from a textbook as a central teaching tool. For these reasons, a great amount of effort has gone into the illustrations. The illustrations are created with four goals in mind:

1. **Completeness** For most figures, it should be possible to understand an experiment or genetic concept by looking at the illustration alone. Students have complained that it is difficult to understand the content of an illustration if they have to keep switching back and forth between the figure and text. In cases where an illustration shows the steps in a scientific process, the steps are described in brief statements that allow the students to understand the whole process (e.g., see Figure 17.11). Likewise, such illustrations should make it easier for instructors to explain these processes in the classroom.
2. **Clarity** The figures have been extensively reviewed by students and instructors. This has helped us to avoid drawing things that may be confusing or unclear. I hope that no one looks at an element in any figure and wonders, "What is that thing?" Aside from being unmistakably drawn, all new elements within each figure are clearly labeled.
3. **Consistency** Before we began to draw the figures, we generated a style sheet that contained recurring elements that are found in many places in the textbook. Examples include the DNA double helix, DNA polymerase, and fruit flies. We agreed on the best way(s) to draw these elements and also what colors they should be. Therefore, as students and instructors progress through this textbook, they become accustomed to the way things should look.
4. **Realism** An important emphasis of this textbook is to make each figure as realistic as possible. When drawing macroscopic elements (e.g., fruit flies, pea plants), the illustrations are based on real images, not on cartoonlike simplifications. Our most challenging goal, and one that we feel has been achieved most successfully, is the realism of our molecular drawings. Whenever possible, we have tried to depict molecular elements according to their actual structures, if such structures are known. For example, the ways we have drawn RNA polymerase, DNA polymerase, DNA helicase, and ribosomes are based on their crystal structures. When a

student sees a figure in this textbook that illustrates an event in transcription, RNA polymerase is depicted in a way that is as realistic as possible (e.g., Figure 14.8, see below).



Key points:

- RNA polymerase slides along the DNA, creating an open complex as it moves.
- The DNA strand known as the template strand is used to make a complementary copy of RNA as an RNA–DNA hybrid.
- RNA polymerase moves along the template strand in a 3′ to 5′ direction, and RNA is synthesized in a 5′ to 3′ direction using nucleoside triphosphates as precursors. Pyrophosphate is released (not shown).
- The complementarity rule is the same as the AT/GC rule except that U is substituted for T in the RNA.

WRITING STYLE

Motivation in learning often stems from enjoyment. If you enjoy what you're reading, you are more likely to spend longer amounts of time with it and focus your attention more crisply. The writing style of this book is meant to be interesting, down to Earth, and easy to follow. Each section of every chapter begins with an overview of the contents of that section, usually with a table or figure that summarizes the broad points. The section then examines how those broad points were discovered experimentally, as well as explaining many of the finer scientific details. Important terms are introduced in a boldface font. These terms are also found at the end of the chapter and in the glossary.

There are various ways to make a genetics book interesting and inspiring. The subject matter itself is pretty amazing, so it's not difficult to build on that. In addition to describing the concepts and experiments in ways that motivate students, it is important to draw on examples that bring the concepts to life. In a genetics book, many of these examples come from the medical realm. This textbook contains lots of examples of human diseases that exemplify some of the underlying principles of genetics. Students often say they remember certain genetic concepts because they remember how defects in certain genes can cause disease. For example, defects in DNA repair genes cause a higher predisposition to develop cancer. In addition, I have tried to be evenhanded in providing examples from the microbial and plant world. Finally, students are often interested in applications of genetics that affect their everyday lives. Because we frequently hear about genetics in the news, it's inspiring for students to learn the underlying basis for such technologies. Chapters 19 to 22 are devoted to genetic technologies, and applications of these and other technologies are found throughout this textbook. By the end of their genetics course, students should come away with a greater appreciation for the influence of genetics in their lives.

SUGGESTIONS WELCOME!

It seems very appropriate to use the word *evolution* to describe the continued development of this textbook. I welcome any and all comments. The refinement of any science textbook requires input from instructors and their students. These include comments regarding writing, illustrations, supplements, factual content, and topics that may need greater or less emphasis. You are invited to contact me at:

Dr. Rob Brooker
Dept. of Genetics, Cell Biology, and Development
University of Minnesota
6-160 Jackson Hall
321 Church St.
Minneapolis, MN 55455
brook005@umn.edu

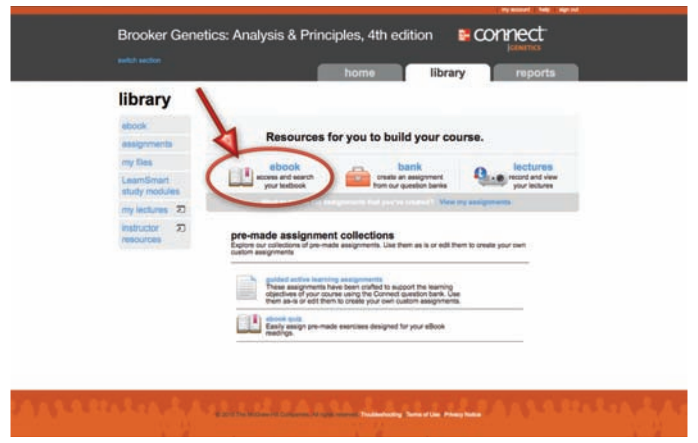
TEACHING AND LEARNING SUPPLEMENTS

www.mhhe.com/brookerconcepts

McGraw-Hill Connect™ Genetics provides online presentation, assignment, and assessment solutions. It connects your students with the tools and resources they'll need to achieve success.

With Connect™ Genetics you can deliver assignments, quizzes, and tests online. A set of questions and activities are presented for every chapter. As an instructor, you can edit existing questions and author entirely new problems. Track individual student performance—by question, assignment, or in relation to the class overall—with detailed grade reports. Integrate grade reports easily with Learning Management Systems (LMS), such as Blackboard® and WebCT. And much more.

ConnectPlus™ Genetics provides students with all the advantages of Connect™ Genetics, plus 24/7 online access to an ebook. To learn more visit www.mcgrawhillconnect.com


Brooker Genetics: Analysis & Principles, 4th edition

Resources for you to build your course.

pre-made assignment collections

connect
home
library
reports
library
ebook
assignments
my files
LearnSmart study modules
my lectures
instructor resources
ebook
access and search your textbook
bank
create an assignment from our question banks
lectures
record and view your lectures

PRESENTATION CENTER

Build instructional materials wherever, whenever, and however you want!

www.mhhe.com/brookerconcepts

The Presentation Center is an online digital library containing photos, artwork, animations, and other media tools that can be used to create customized lectures, visually enhanced tests and quizzes, compelling course websites, or attractive printed support materials. All assets are copyrighted by McGraw-Hill Higher Education, but can be used by instructors for classroom purposes. The visual resources in this collection include

- **FlexArt Image PowerPoints®** Full-color digital files of all illustrations in the book with editable labels can be readily incorporated into lecture presentations, exams, or custom-made classroom materials. All files are preinserted into PowerPoint slides for ease of lecture preparation.
- **Photos** The photo collection contains digital files of photographs from the text, which can be reproduced for multiple classroom uses.
- **Tables** Every table that appears in the text has been saved in electronic form for use in classroom presentations or quizzes.
- **Animations** Numerous full-color animations illustrating important processes are also provided. Harness the visual effect of concepts in motion by importing these files into classroom presentations or online course materials.
- **PowerPoint Lecture Outlines** Ready-made presentations that combine art and lecture notes are provided for each chapter of the text.
- **PowerPoint Slides** For instructors who prefer to create their lectures from scratch, all illustrations, photos, tables and animations are preinserted by chapter into blank PowerPoint slides.

FOR THE STUDENT

Student Study Guide/Solutions Manual Online

The Study Guide follows the order of sections and subsections in the textbook and summarizes the main points in the text, figures, and tables. It also contains concept-building exercises, self-help quizzes, and practice exams. The solutions to the end-of-chapter problems and questions aid the students in developing their problem-solving skills by providing the steps for each solution.

Companion Website

www.mhhe.com/brookerconcepts

The Brooker *Concepts of Genetics* companion website offers an extensive array of learning tools, including a variety of quizzes for each chapter, interactive genetics problems, animations and more.

McGraw-Hill ConnectPlus™ interactive learning platform provides all of the benefits of Connect: online presentation tools, auto-grade assessments, and powerful reporting—all in an easy-to-use interface, as well as a customizable, assignable ebook. This media-rich version of the book is available through the McGraw-Hill Connect™platform and allows seamless integration of text, media, and assessment.

By choosing ConnectPlus™, instructors are providing their students with a powerful tool for improving academic performance and truly mastering course material. ConnectPlus™ allows students to practice important skills at their own pace and on their own schedule. Students' assessment results and instructors' feedback are saved online—so students can continually review their progress and plot their course to success. Learn more at:

www.mcgrawhillconnect.com

BREAKING NEWS! McGraw-Hill, Blackboard Announce New Partnership!

Create assignments and assessments from the library or anything you find on the Web.

SELECT ANY OF THE 6 SUBJECT AREAS (e.g. BUSINESS AND ECONOMICS, CAREER)

connect
SIGN IN
HOME
ABOUT
SUBJECTS
PRODUCT TOUR
SUPPORT
Learn More...
Bb
LESS MANAGING.
MORE TEACHING.
GREATER LEARNING.
Introduction 101
connect
library
reports
BUSINESS AND ECONOMICS
CAREER
HUMANITIES AND SOCIAL SCIENCES
PHYSICAL AND EARTH SCIENCES
ENGINEERING

McGraw-Hill Higher Education and Blackboard have teamed up.

Blackboard, the Web-based course-management system, has partnered with McGraw-Hill to better allow students and faculty to use online materials and activities to complement face-to-face teaching. Blackboard features exciting social learning and teaching tools that foster more logical, visually impactful and active learning opportunities for students. You'll transform your closed-door classrooms into communities where students remain connected to their educational experience 24 hours a day.

This partnership allows you and your students access to McGraw-Hill's Connect™ and Create™ right from within your Blackboard course—all with one single sign-on.

Not only do you get single sign-on with Connect™ and Create™, you also get deep integration of McGraw-Hill content and content engines right in Blackboard. Whether you're choosing a book for your course or building Connect™ assignments, all the tools you need are right where you want them—inside of Blackboard.

Gradebooks are now seamless. When a student completes an integrated Connect™ assignment, the grade for that assignment automatically (and instantly) feeds your Blackboard grade center.

McGraw-Hill and Blackboard can now offer you easy access to industry leading technology and content, whether your campus hosts it, or we do. Be sure to ask your local McGraw-Hill representative for details.

ACKNOWLEDGMENTS

The production of a textbook is truly a collaborative effort, and I am greatly indebted to a variety of people. This textbook has gone through multiple rounds of rigorous revision that involved the input of faculty, students, editors, and educational and media specialists. Their collective contributions are reflected in the final outcome.

Let me begin by acknowledging the many people at McGraw-Hill whose efforts are amazing. My highest praise goes to Mandy Clark (Developmental Editor) and Elizabeth Sievers (Director of Development), who managed many aspects of this project. I also would like to thank Janice Roerig-Blong (Publisher) for her patience in overseeing this project. She has the unenviable job of managing the budget for the book and that is not an easy task. Other people at McGraw-Hill have played key roles in producing an actual book and the supplements that go along with it. In particular, Jayne Klein (Project Manager) has done a superb job of managing the components that need to be assembled to produce a book, along with Sherry Kane (Buyer). I would also like to thank John Leland (Photo Research Coordinator), who acted as an interface between me and the photo company. In addition, my gratitude goes to David Hash (Designer), who provided much input into the internal design of the book as well as creating an awesome cover. Finally, I would like to thank Patrick Reidy (Marketing Manager), whose major efforts begin when the first edition comes out!

With regard to the content of the book, Joni Fraser (Freelance Developmental Editor) has worked closely with me in developing a book that is clear, consistent, and easy for students to follow. She analyzed all of the chapters in the textbook and made improvements with regard to content, art, and organization. She also scrutinized the text for clarity and logic. I would also like to thank Linda Davoli (Freelance Copy Editor) for making grammatical improvements throughout the text and art, which has significantly improved the text's clarity.

I would also like to extend my thanks to Bonnie Briggle and everyone at Lachina Publishing Services, including the many artists who have played important roles in developing the art for this textbook. Also, the folks at Lachina Publishing Services worked with great care in the paging of the book, making sure that the figures and relevant text are as close to each other as possible. Likewise, the people at Pronk & Associates have done a great job of locating many of the photographs that have been used in this textbook.

Finally, I want to thank the many scientists who reviewed the chapters of this textbook. Their broad insights and constructive suggestions were an important factor that shaped its final content and organization. I am truly grateful for their time and effort.

REVIEWERS

Preston Aldrich, *Benedictine University*
Diya Banerjee, *Virginia Tech University*
Vernon W. Bauer, *Francis Marion University*
Mark Brick, *Colorado State University*
Aaron Cassill, *University of Texas at San Antonio*
Bruce Chase, *University of Nebraska at Omaha*
Erin Cram, *Northeastern University*
Sandra L. Davis, *University of Indianapolis*
Steve Denison, *Eckerd College*
Michele Engel, *University of Colorado–Denver*
Jayant Ghiara, *University of California–San Diego*
Meredith Hamilton, *Oklahoma State University*
Stephen C. Hedman, *University of Minnesota*
Robert Hinrichsen, *Indiana University of Pennsylvania*
David Kass, *Eastern Michigan University*
Ekaterina N. Kaverina, *East Tennessee State University*
Sarah Kenick, *University of New Hampshire*
Michael Kielb, *Eastern Michigan University*
Brian Kreiser, *University of Southern Mississippi*
Michael Lehmann, *University of Arkansas*
Haiying Liang, *Clemson University*

Shawn Meagher, *Western Illinois University*
Marcie Moehnke, *Baylor University*
Roderick M. Morgan, *Grand Valley State University*
James Morris, *Clemson University*
Sang-Chul Nam, *Baylor University, Waco, TX*
John C. Osterman, *University of Nebraska–Lincoln*
William A. Rosche, *Richard Stockton College of New Jersey*
Lara Soowal, *University of California–San Diego*
Tzvi Tzfira, *University of Michigan*
Timothy Walston, *Truman State University*
Yunqiu Wang, *University of Miami*
Cindy White, *University of Northern Colorado*
Malcolm Zellars, *Georgia State University*
Jianzhi Zhang, *University of Michigan*

ACCURACY CHECKERS

Vernon W. Bauer, *Francis Marion University*
Mark Brick, *Colorado State University*
Aaron Cassill, *University of Texas at San Antonio*
Bruce Chase, *University of Nebraska at Omaha*
Erin Cram, *Northeastern University*
Sandra L. Davis, *University of Indianapolis*
Steve Denison, *Eckerd College*
Michele Engel, *University of Colorado Denver*
Stephen C. Hedman, *University of Minnesota*
David Kass, *Eastern Michigan University*
Sarah Kenick, *University of New Hampshire*
Michael Lehmann, *University of Arkansas*
Shawn Meagher, *Western Illinois University*
Marcie Moehnke, *Baylor University*
Roderick M. Morgan, *Grand Valley State University*
James Morris, *Clemson University*
Sang-Chul Nam, *Baylor University, Waco, TX*
John C. Osterman, *University of Nebraska–Lincoln*
Tzvi Tzfira, *University of Michigan*
Yunqiu Wang, *University of Miami*
Tim Walston, *Truman State University*
Malcolm Zellars, *Georgia State University*
Jianzhi Zhang, *University of Michigan*

ABOUT THE AUTHOR

Rob Brooker is a professor in the Department of Genetics, Cell Biology, and Development at the University of Minnesota–Minneapolis. He received his B.A. in biology from Wittenberg University in 1978 and his Ph.D. in genetics from Yale University in 1983. At Harvard, he conducted postdoctoral studies on the lactose permease, which is the product of the *lacY* gene of the *lac* operon. He continues his work on transporters at the University of Minnesota. Dr. Brooker's laboratory primarily investigates the structure, function, and regulation of iron transporters found in bacteria and *C. elegans*. At the University of Minnesota he teaches undergraduate courses in biology, genetics, and cell biology.

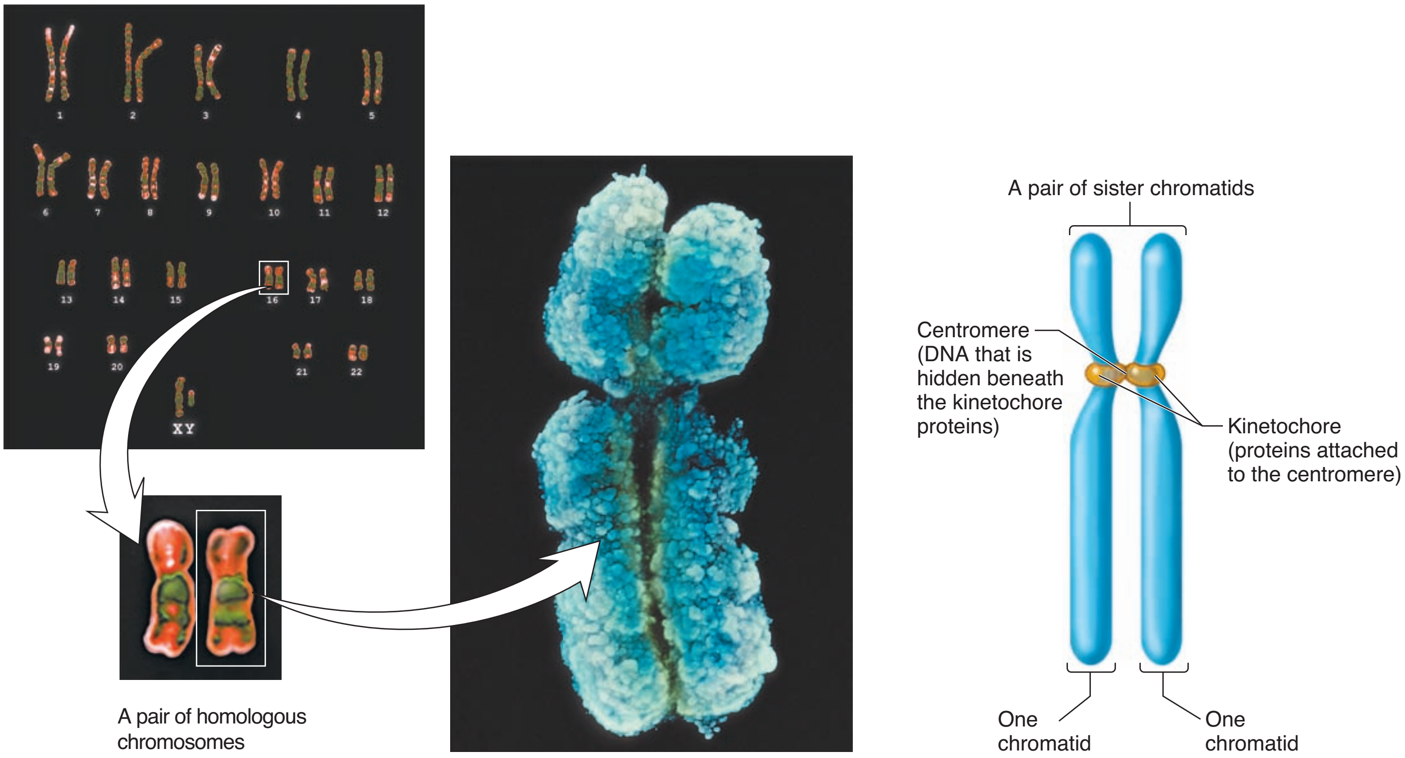
Dedication

To my wife, Deborah, and our children,
Daniel, Nathan, and Sarah

A Visual Guide to
CONCEPTS OF GENETICS

Instructional Art

Brooker's Concepts of Genetics brings key concepts to life with its unique style of illustration.

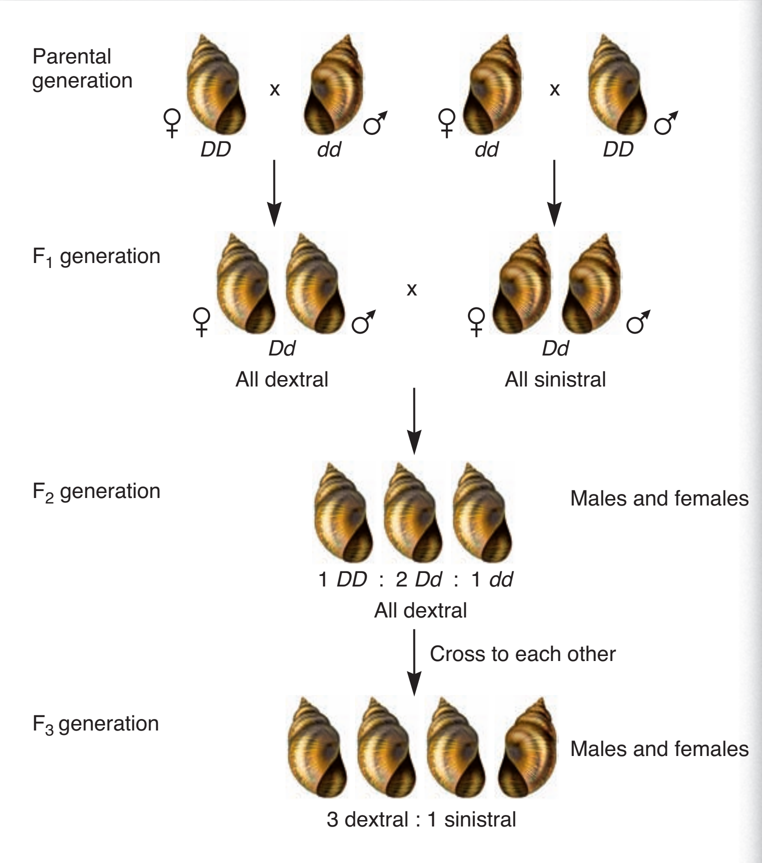

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
XY
A pair of homologous chromosomes
A pair of sister chromatids
Centromere (DNA that is hidden beneath the kinetochore proteins)
Kinetochore (proteins attached to the centromere)
One chromatid
One chromatid

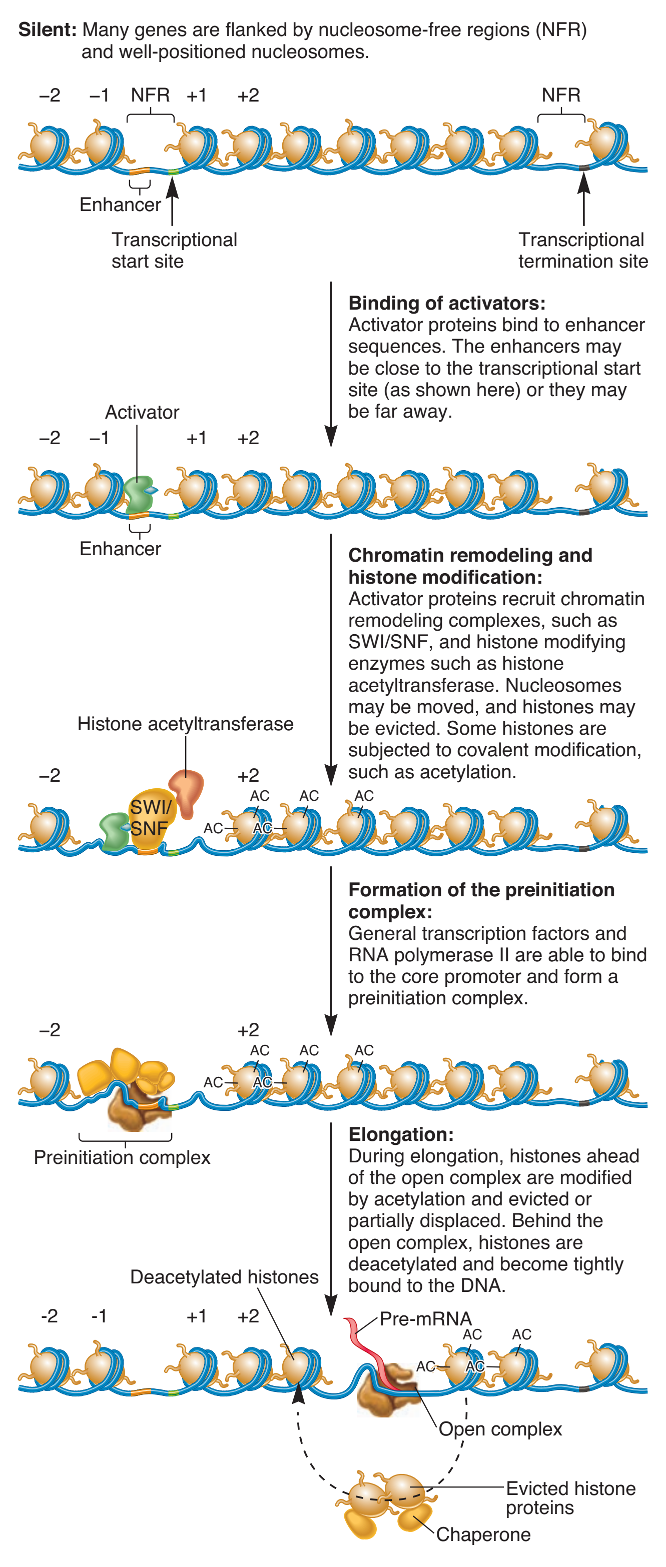
(a) Homologous chromosomes and sister chromatids

(b) Schematic drawing of sister chromatids

The digitally rendered images have a vivid three-dimensional look that will stimulate a student's interest and enthusiasm.

Each figure is carefully designed to follow closely with the text material.


Parental generation
♀
DD
x
dd
♂
♀
dd
x
DD
♂
F$_1$ generation
♀
Dd
♂
All dextral
x
♀
Dd
♂
All sinistral
F$_2$ generation
Males and females
1 DD : 2 Dd : 1 dd
All dextral
Cross to each other
F$_3$ generation
Males and females
3 dextral : 1 sinistral


Silent: Many genes are flanked by nucleosome-free regions (NFR) and well-positioned nucleosomes.

−2
−1
NFR
+1
+2
NFR
Enhancer
Transcriptional start site
Transcriptional termination site

Binding of activators: Activator proteins bind to enhancer sequences. The enhancers may be close to the transcriptional start site (as shown here) or they may be far away.

Activator
−2
−1
+1
+2
Enhancer

Chromatin remodeling and histone modification: Activator proteins recruit chromatin remodeling complexes, such as SWI/SNF, and histone modifying enzymes such as histone acetyltransferase. Nucleosomes may be moved, and histones may be evicted. Some histones are subjected to covalent modification, such as acetylation.

Histone acetyltransferase
−2
+2
SWI/
SNF
AC
AC
AC
AC
AC

Formation of the preinitiation complex: General transcription factors and RNA polymerase II are able to bind to the core promoter and form a preinitiation complex.

−2
+2
AC
AC
AC
AC
AC
Preinitiation complex

Elongation: During elongation, histones ahead of the open complex are modified by acetylation and evicted or partially displaced. Behind the open complex, histones are deacetylated and become tightly bound to the DNA.

Deacetylated histones
-2
-1
+1
+2
Pre-mRNA
AC
AC
AC
AC
Open complex
Evicted histone proteins
Chaperone

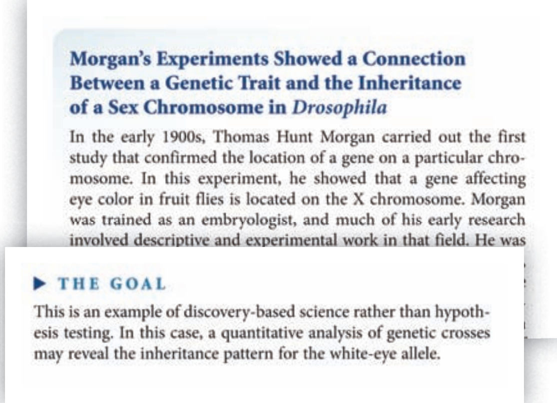
Every illustration was drawn with four goals in mind: completeness, clarity, consistency, and realism.

Learning Through Experimentation

Many chapters contain an experiment that is presented according to the scientific method. These experiments are integrated within the chapters and flow with the rest of the textbook. As you read the experiments, which can be hypothesis-testing or discovery-based science, you will simultaneously explore the scientific method and the genetic principles learned from this approach.

BACKGROUND OBSERVATIONS

Each experiment begins with a description of the information that led researchers to study a hypothesis-driven or discovery-based problem. Detailed information about the researchers and the experimental challenges they faced help students to understand actual research.



Morgan's Experiments Showed a Connection Between a Genetic Trait and the Inheritance of a Sex Chromosome in *Drosophila*

In the early 1900s, Thomas Hunt Morgan carried out the first study that confirmed the location of a gene on a particular chromosome. In this experiment, he showed that a gene affecting eye color in fruit flies is located on the X chromosome. Morgan was trained as an embryologist, and much of his early research involved descriptive and experimental work in that field. He was

▶ **THE GOAL**

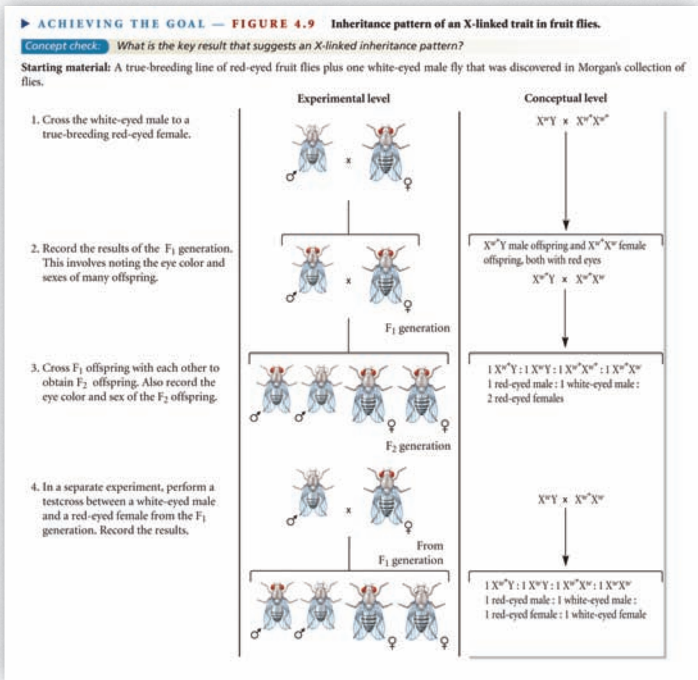
This is an example of discovery-based science rather than hypothesis testing. In this case, a quantitative analysis of genetic crosses may reveal the inheritance pattern for the white-eye allele.

THE HYPOTHESIS OR THE GOAL

The student is given a possible explanation for the observed phenomenon that will be tested or the question researchers were hoping to answer. This section reinforces the scientific method and allows students to experience the process for themselves.

TESTING THE HYPOTHESIS OR ACHIEVING THE GOAL

This section illustrates the experimental process, including the actual steps followed by scientists to test their hypothesis or study a question. Science comes alive for students with this detailed look at experimentation.

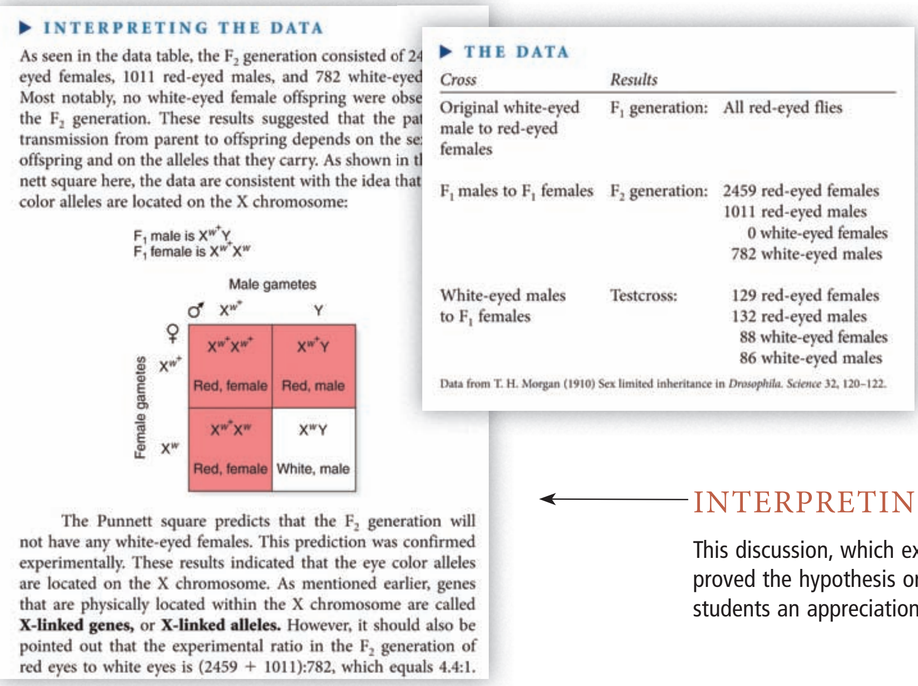


▶ **ACHIEVING THE GOAL — FIGURE 4.9** **Inheritance pattern of an X-linked trait in fruit flies.**

Concept check: *What is the key result that suggests an X-linked inheritance pattern?*

Starting material: A true-breeding line of red-eyed fruit flies plus one white-eyed male fly that was discovered in Morgan's collection of flies.

	Experimental level	Conceptual level
1. Cross the white-eyed male to a true-breeding red-eyed female.		$X^wY \times X^{w^+}X^{w^+}$
2. Record the results of the F_1 generation. This involves noting the eye color and sexes of many offspring.	F_1 generation	$X^{w^+}Y$ male offspring and $X^{w^+}X^w$ female offspring, both with red eyes $X^{w^+}Y \times X^{w^+}X^w$
3. Cross F_1 offspring with each other to obtain F_2 offspring. Also record the eye color and sex of the F_2 offspring.	F_2 generation	$1\ X^{w^+}Y : 1\ X^wY : 1\ X^{w^+}X^{w^+} : 1\ X^{w^+}X^w$ 1 red-eyed male : 1 white-eyed male : 2 red-eyed females
4. In a separate experiment, perform a testcross between a white-eyed male and a red-eyed female from the F_1 generation. Record the results.	From F_1 generation	$X^wY \times X^{w^+}X^w$ $1\ X^{w^+}Y : 1\ X^wY : 1\ X^{w^+}X^w : 1\ X^wX^w$ 1 red-eyed male : 1 white-eyed male : 1 red-eyed female : 1 white-eyed female



▶ **INTERPRETING THE DATA**

As seen in the data table, the F_2 generation consisted of 24
eyed females, 1011 red-eyed males, and 782 white-eyed
Most notably, no white-eyed female offspring were obse
the F_2 generation. These results suggested that the pat
transmission from parent to offspring depends on the se
offspring and on the alleles that they carry. As shown in tl
nett square here, the data are consistent with the idea that
color alleles are located on the X chromosome:

F_1 male is $X^{w^+}Y$
F_1 female is $X^{w^+}X^w$

Male gametes

Female gametes	♂ X^{w^+}	Y
♀ X^{w^+}	$X^{w^+}X^{w^+}$ Red, female	$X^{w^+}Y$ Red, male
X^w	$X^{w^+}X^w$ Red, female	X^wY White, male

The Punnett square predicts that the F_2 generation will not have any white-eyed females. This prediction was confirmed experimentally. These results indicated that the eye color alleles are located on the X chromosome. As mentioned earlier, genes that are physically located within the X chromosome are called **X-linked genes,** or **X-linked alleles.** However, it should also be pointed out that the experimental ratio in the F_2 generation of red eyes to white eyes is (2459 + 1011):782, which equals 4.4:1.

▶ **THE DATA**

Cross	*Results*	
Original white-eyed male to red-eyed females	F_1 generation:	All red-eyed flies
F_1 males to F_1 females	F_2 generation:	2459 red-eyed females 1011 red-eyed males 0 white-eyed females 782 white-eyed males
White-eyed males to F_1 females	Testcross:	129 red-eyed females 132 red-eyed males 88 white-eyed females 86 white-eyed males

Data from T. H. Morgan (1910) Sex limited inheritance in *Drosophila. Science* 32, 120–122.

THE DATA

Actual data from the original research paper help students understand how real-life research results are reported. Each experiment's results are discussed in the context of the larger genetic principle to help students understand the implications and importance of the research.

INTERPRETING THE DATA

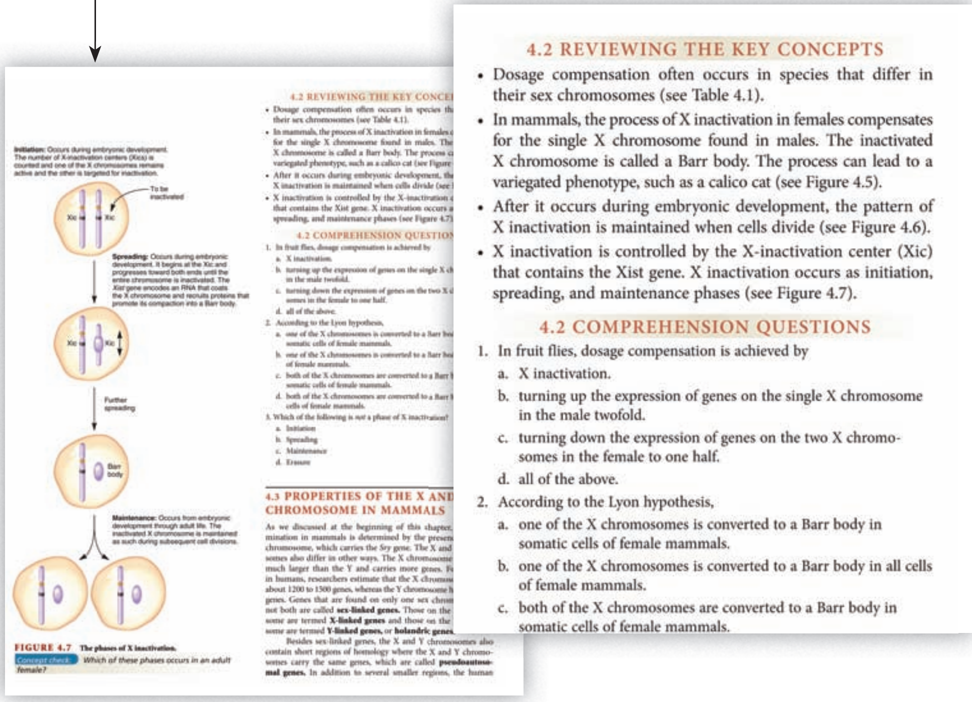
This discussion, which examines whether the experimental data supported or disproved the hypothesis or provided new information to propose a hypothesis, gives students an appreciation for scientific interpretation.

Supportive Features and Materials Throughout the Chapter

These study tools and problems are crafted to aid students in reviewing key information in the text and developing a wide range of skills. They also develop a student's cognitive, writing, analytical, computational, and collaborative abilities.

CONCEPT CHECK QUESTIONS

Students can test their knowledge and understanding with Concept check *questions that are associated with the figure legends. These questions often go beyond simple recall of information and ask students to apply or interpret information presented in the illustrations.*



Initiation: Occurs during embryonic development. The number of X-inactivation centers (Xics) is counted and one of the X chromosomes remains active and the other is targeted for inactivation.
To be inactivated
Xic
Xic
Spreading: Occurs during embryonic development. It begins at the Xic and progresses toward both ends until the entire chromosome is inactivated. The Xist gene encodes an RNA that coats the X chromosome and recruits proteins that promote its compaction into a Barr body.
Further spreading
Barr body
Maintenance: Occurs from embryonic development through adult life. The inactivated X chromosome is maintained as such during subsequent cell divisions.

FIGURE 4.7 The phases of X inactivation.
Concept check: *Which of these phases occurs in an adult female?*

4.2 REVIEWING THE KEY CONCEPTS

- Dosage compensation often occurs in species that differ in their sex chromosomes (see Table 4.1).
- In mammals, the process of X inactivation in females compensates for the single X chromosome found in males. The inactivated X chromosome is called a Barr body. The process can lead to a variegated phenotype, such as a calico cat (see Figure 4.5).
- After it occurs during embryonic development, the pattern of X inactivation is maintained when cells divide (see Figure 4.6).
- X inactivation is controlled by the X-inactivation center (Xic) that contains the Xist gene. X inactivation occurs as initiation, spreading, and maintenance phases (see Figure 4.7).

4.2 COMPREHENSION QUESTIONS

1. In fruit flies, dosage compensation is achieved by
   a. X inactivation.
   b. turning up the expression of genes on the single X chromosome in the male twofold.
   c. turning down the expression of genes on the two X chromosomes in the female to one half.
   d. all of the above.
2. According to the Lyon hypothesis,
   a. one of the X chromosomes is converted to a Barr body in somatic cells of female mammals.
   b. one of the X chromosomes is converted to a Barr body in all cells of female mammals.
   c. both of the X chromosomes are converted to a Barr body in somatic cells of female mammals.

REVIEWING THE KEY CONCEPTS

These bulleted lists at the end of each section help students identify important concepts. Students should understand these concepts before moving on to the next section.

COMPREHENSION QUESTIONS

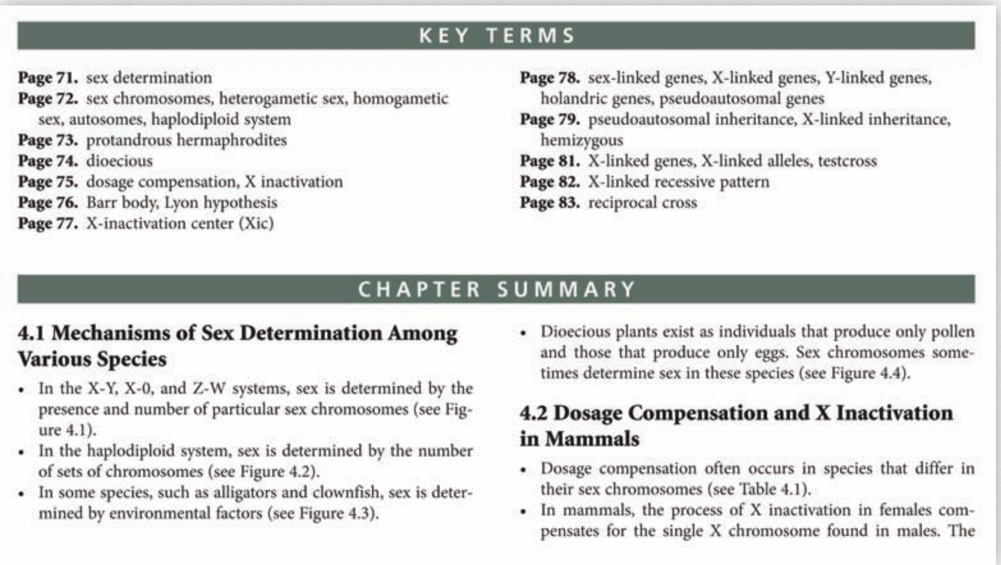
Multiple choice questions found at the end of each section allow students an opportunity to test their knowledge of key information and concepts. This helps students better identify what they know and don't know, before tackling more concepts.

KEY TERMS

Providing the key terms from the chapter enhances student development of vital vocabulary necessary for the understanding and application of chapter content. Important terms are boldfaced throughout the chapter and page referenced at the end of each chapter for reflective study.

CHAPTER SUMMARY

These bulleted summaries organized by section emphasize the main concepts of the chapter to provide students with a thorough review of the main topics covered.



KEY TERMS

Page 71. sex determination
Page 72. sex chromosomes, heterogametic sex, homogametic sex, autosomes, haplodiploid system
Page 73. protandrous hermaphrodites
Page 74. dioecious
Page 75. dosage compensation, X inactivation
Page 76. Barr body, Lyon hypothesis
Page 77. X-inactivation center (Xic)
Page 78. sex-linked genes, X-linked genes, Y-linked genes, holandric genes, pseudoautosomal genes
Page 79. pseudoautosomal inheritance, X-linked inheritance, hemizygous
Page 81. X-linked genes, X-linked alleles, testcross
Page 82. X-linked recessive pattern
Page 83. reciprocal cross

CHAPTER SUMMARY

4.1 Mechanisms of Sex Determination Among Various Species

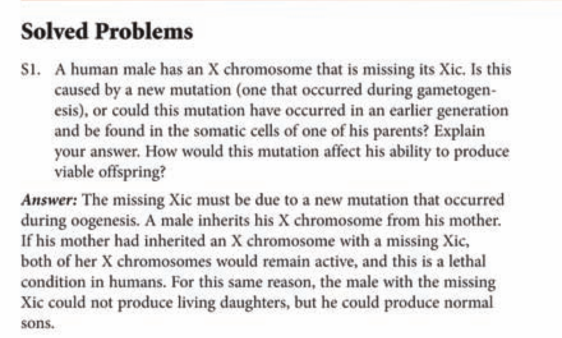
- In the X-Y, X-0, and Z-W systems, sex is determined by the presence and number of particular sex chromosomes (see Figure 4.1).
- In the haplodiploid system, sex is determined by the number of sets of chromosomes (see Figure 4.2).
- In some species, such as alligators and clownfish, sex is determined by environmental factors (see Figure 4.3).
- Dioecious plants exist as individuals that produce only pollen and those that produce only eggs. Sex chromosomes sometimes determine sex in these species (see Figure 4.4).

4.2 Dosage Compensation and X Inactivation in Mammals

- Dosage compensation often occurs in species that differ in their sex chromosomes (see Table 4.1).
- In mammals, the process of X inactivation in females compensates for the single X chromosome found in males. The

SOLVED PROBLEMS

These problems walk students through the solutions, allowing them to see the steps involved in solving the problems. These provide a reference for when students encounter similar problems later.



PROBLEM SETS & INSIGHTS

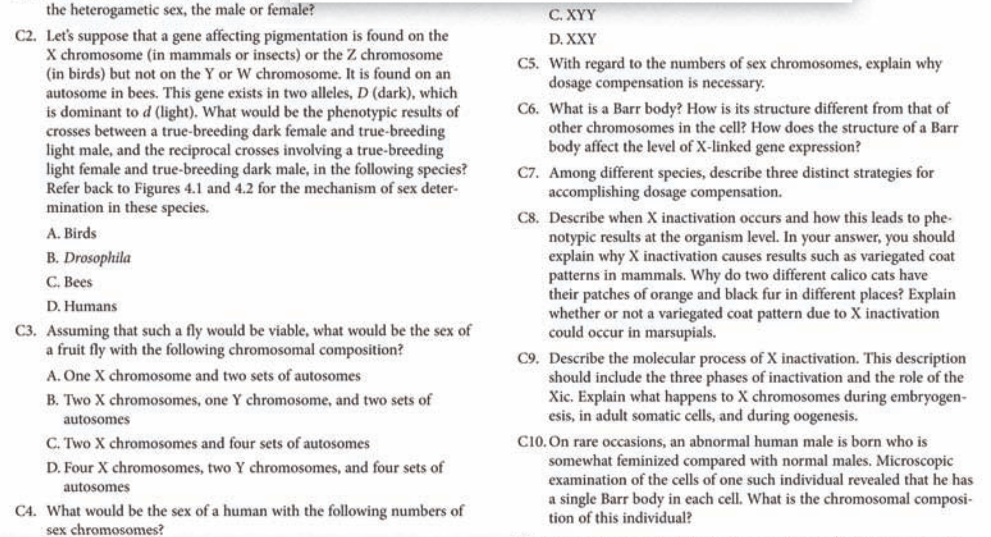
Solved Problems

S1. A human male has an X chromosome that is missing its Xic. Is this caused by a new mutation (one that occurred during gametogenesis), or could this mutation have occurred in an earlier generation and be found in the somatic cells of one of his parents? Explain your answer. How would this mutation affect his ability to produce viable offspring?

Answer: The missing Xic must be due to a new mutation that occurred during oogenesis. A male inherits his X chromosome from his mother. If his mother had inherited an X chromosome with a missing Xic, both of her X chromosomes would remain active, and this is a lethal condition in humans. For this same reason, the male with the missing Xic could not produce living daughters, but he could produce normal sons.

CONCEPTUAL QUESTIONS

These questions test the understanding of basic genetic principles. The student is given many questions with a wide range of difficulty. Some require critical thinking skills, and some require the student to write coherent answers in an essay form.



Conceptual Questions

C1. In the X-Y and Z-W systems of sex de the heterogametic sex, the male or female?

C2. Let's suppose that a gene affecting pigmentation is found on the X chromosome (in mammals or insects) or the Z chromosome (in birds) but not on the Y or W chromosome. It is found on an autosome in bees. This gene exists in two alleles, *D* (dark), which is dominant to *d* (light). What would be the phenotypic results of crosses between a true-breeding dark female and true-breeding light male, and the reciprocal crosses involving a true-breeding light female and true-breeding dark male, in the following species? Refer back to Figures 4.1 and 4.2 for the mechanism of sex determination in these species.

A. Birds
B. *Drosophila*
C. Bees
D. Humans

C3. Assuming that such a fly would be viable, what would be the sex of a fruit fly with the following chromosomal composition?

A. One X chromosome and two sets of autosomes
B. Two X chromosomes, one Y chromosome, and two sets of autosomes
C. Two X chromosomes and four sets of autosomes
D. Four X chromosomes, two Y chromosomes, and four sets of autosomes

C4. What would be the sex of a human with the following numbers of sex chromosomes?

C. XYY
D. XXY

C5. With regard to the numbers of sex chromosomes, explain why dosage compensation is necessary.

C6. What is a Barr body? How is its structure different from that of other chromosomes in the cell? How does the structure of a Barr body affect the level of X-linked gene expression?

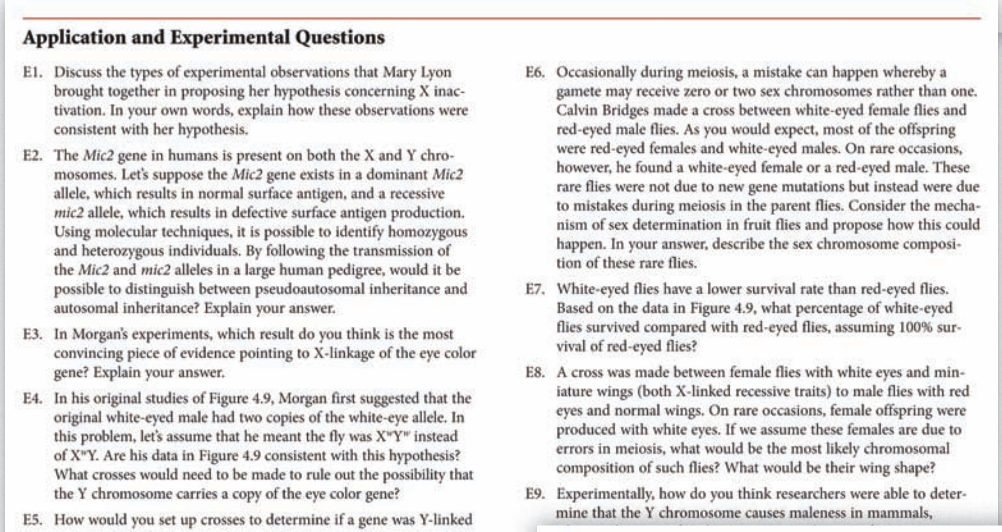
C7. Among different species, describe three distinct strategies for accomplishing dosage compensation.

C8. Describe when X inactivation occurs and how this leads to phenotypic results at the organism level. In your answer, you should explain why X inactivation causes results such as variegated coat patterns in mammals. Why do two different calico cats have their patches of orange and black fur in different places? Explain whether or not a variegated coat pattern due to X inactivation could occur in marsupials.

C9. Describe the molecular process of X inactivation. This description should include the three phases of inactivation and the role of the Xic. Explain what happens to X chromosomes during embryogenesis, in adult somatic cells, and during oogenesis.

C10. On rare occasions, an abnormal human male is born who is somewhat feminized compared with normal males. Microscopic examination of the cells of one such individual revealed that he has a single Barr body in each cell. What is the chromosomal composition of this individual?

C11. How many Barr bodies would you expect to find in humans with the following abnormal compositions of sex chromosomes?



Application and Experimental Questions

E1. Discuss the types of experimental observations that Mary Lyon brought together in proposing her hypothesis concerning X inactivation. In your own words, explain how these observations were consistent with her hypothesis.

E2. The *Mic2* gene in humans is present on both the X and Y chromosomes. Let's suppose the *Mic2* gene exists in a dominant *Mic2* allele, which results in normal surface antigen, and a recessive *mic2* allele, which results in defective surface antigen production. Using molecular techniques, it is possible to identify homozygous and heterozygous individuals. By following the transmission of the *Mic2* and *mic2* alleles in a large human pedigree, would it be possible to distinguish between pseudoautosomal inheritance and autosomal inheritance? Explain your answer.

E3. In Morgan's experiments, which result do you think is the most convincing piece of evidence pointing to X-linkage of the eye color gene? Explain your answer.

E4. In his original studies of Figure 4.9, Morgan first suggested that the original white-eyed male had two copies of the white-eye allele. In this problem, let's assume that he meant the fly was X^wY^w instead of X^wY. Are his data in Figure 4.9 consistent with this hypothesis? What crosses would need to be made to rule out the possibility that the Y chromosome carries a copy of the eye color gene?

E5. How would you set up crosses to determine if a gene was Y-linked versus X-linked?

E6. Occasionally during meiosis, a mistake can happen whereby a gamete may receive zero or two sex chromosomes rather than one. Calvin Bridges made a cross between white-eyed female flies and red-eyed male flies. As you would expect, most of the offspring were red-eyed females and white-eyed males. On rare occasions, however, he found a white-eyed female or a red-eyed male. These rare flies were not due to new gene mutations but instead were due to mistakes during meiosis in the parent flies. Consider the mechanism of sex determination in fruit flies and propose how this could happen. In your answer, describe the sex chromosome composition of these rare flies.

E7. White-eyed flies have a lower survival rate than red-eyed flies. Based on the data in Figure 4.9, what percentage of white-eyed flies survived compared with red-eyed flies, assuming 100% survival of red-eyed flies?

E8. A cross was made between female flies with white eyes and miniature wings (both X-linked recessive traits) to male flies with red eyes and normal wings. On rare occasions, female offspring were produced with white eyes. If we assume these females are due to errors in meiosis, what would be the most likely chromosomal composition of such flies? What would be their wing shape?

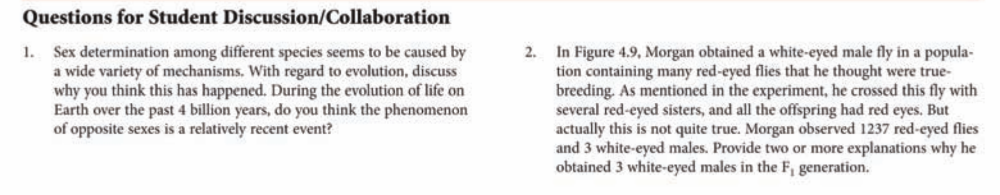
E9. Experimentally, how do you think researchers were able to determine that the Y chromosome causes maleness in mammals,

APPLICATION AND EXPERIMENTAL QUESTIONS

These questions test the ability to analyze data, design experiments, or appreciate the relevance of experimental techniques.

QUESTIONS FOR STUDENT DISCUSSION/ COLLABORATION

These questions encourage students to consider broad concepts and practical problems. Some questions require a substantial amount of computational activities, which can be worked on as a group.

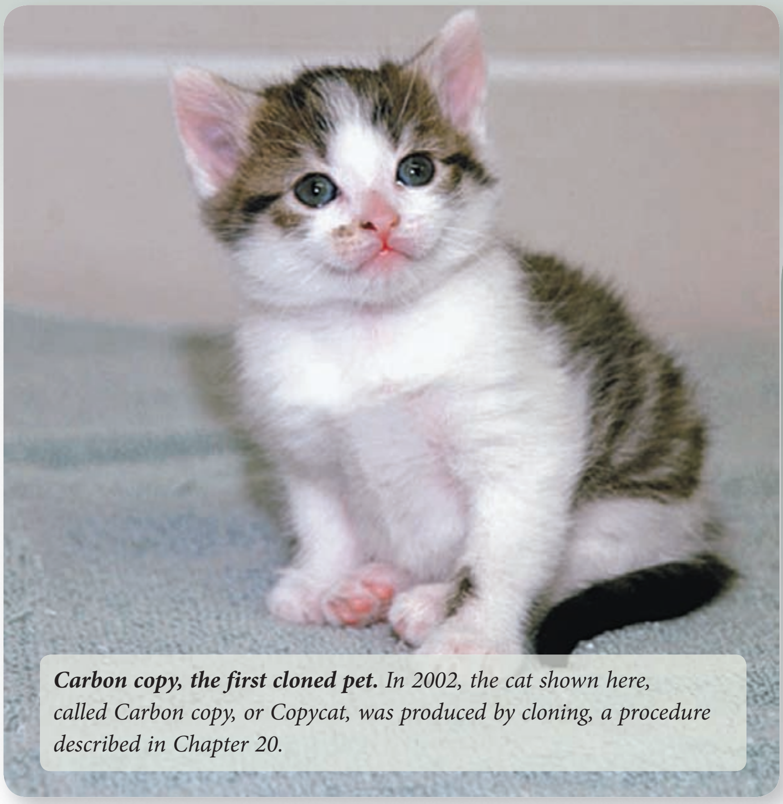


Questions for Student Discussion/Collaboration

1. Sex determination among different species seems to be caused by a wide variety of mechanisms. With regard to evolution, discuss why you think this has happened. During the evolution of life on Earth over the past 4 billion years, do you think the phenomenon of opposite sexes is a relatively recent event?

2. In Figure 4.9, Morgan obtained a white-eyed male fly in a population containing many red-eyed flies that he thought were true-breeding. As mentioned in the experiment, he crossed this fly with several red-eyed sisters, and all the offspring had red eyes. But actually this is not quite true. Morgan observed 1237 red-eyed flies and 3 white-eyed males. Provide two or more explanations why he obtained 3 white-eyed males in the F_1 generation.

PART I INTRODUCTION



Carbon copy, the first cloned pet. *In 2002, the cat shown here, called Carbon copy, or Copycat, was produced by cloning, a procedure described in Chapter 20.*

1

CHAPTER OUTLINE

1.1 The Molecular Expression of Genes

1.2 The Relationship Between Genes and Traits

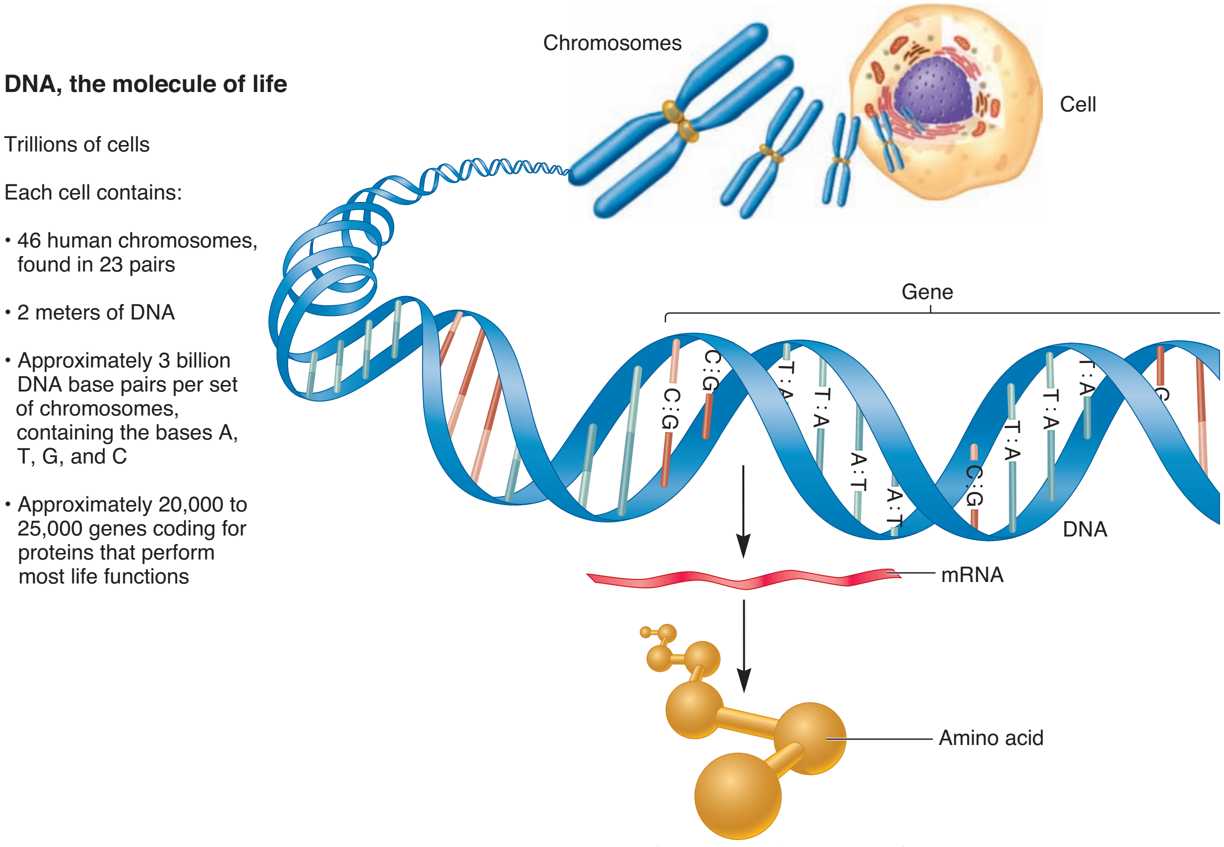
1.3 Fields of Genetics

OVERVIEW OF GENETICS

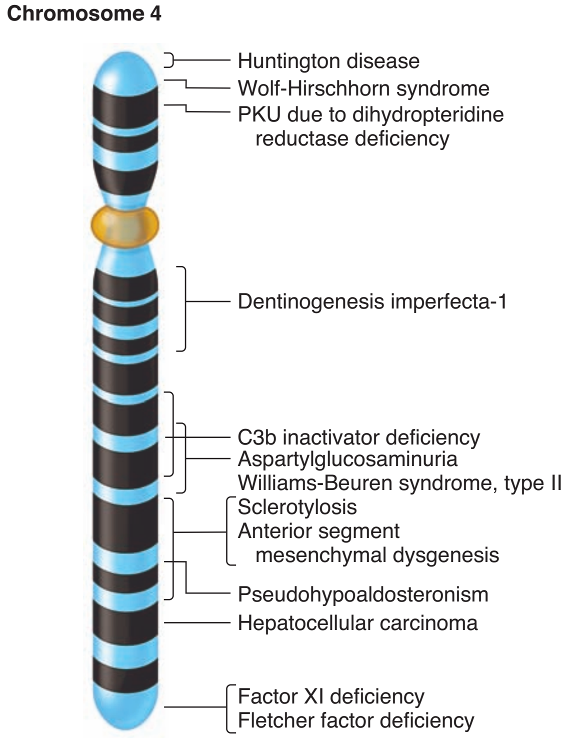
Hardly a week goes by without a major news story involving a genetic breakthrough. The increasing pace of genetic discoveries has become staggering. The Human Genome Project is a case in point. This project began in the United States in 1990, when the National Institutes of Health and the Department of Energy joined forces with international partners to decipher the massive amount of information contained in our **genome**—the **deoxyribonucleic acid (DNA)** found within all of our chromosomes (**Figure 1.1**). Working collectively, a large group of scientists from around the world produced a detailed series of maps that help geneticists navigate through human DNA. Remarkably, in only a decade, they determined the DNA sequence covering over 90% of the human genome. The first draft of this sequence, published in 2001, was nearly 3 billion nucleotide base pairs in length. The completed sequence, published in 2003, has an accuracy greater than 99.99%; fewer than one mistake was made in every 10,000 base pairs (bp)!

Studying the human genome allows us to explore fundamental details about ourselves at the molecular level. The results of the Human Genome Project are expected to shed considerable light on basic questions, such as how many genes we have, how genes direct the activities of living cells, how species evolve, how single cells develop into complex tissues, and how defective genes cause disease. Furthermore, such understanding may lend itself to improvements in modern medicine by providing better diagnoses of diseases and the development of new treatments for them.

As scientists have attempted to unravel the mysteries within our genes, this journey has involved the invention of many new technologies. This textbook emphasizes a large number of these modern approaches. For example, new technologies have made it possible to produce medicines that would otherwise be difficult or impossible to make. An example is human recombinant insulin, sold under the brand name *Humulin,* which is synthesized in strains of *Escherichia coli* bacteria that have been genetically altered by the addition of genes that encode the functional regions of human insulin. The bacteria are grown in a laboratory and make large amounts of human insulin, which is purified and administered to millions of people with insulin-dependent diabetes. Chapter 20 describes the production of insulin in greater detail and also examines other ways that genetic approaches have applications in the area of biotechnology.


DNA, the molecule of life
Trillions of cells
Each cell contains:
• 46 human chromosomes, found in 23 pairs
• 2 meters of DNA
• Approximately 3 billion DNA base pairs per set of chromosomes, containing the bases A, T, G, and C
• Approximately 20,000 to 25,000 genes coding for proteins that perform most life functions
Chromosomes
Cell
Gene
C:G
C:G
T:A
T:A
A:T
A:T
C:G
T:A
T:A
T:A
DNA
mRNA
Amino acid
Protein (composed of amino acids)

(a) The genetic composition of humans


Chromosome 4
Huntington disease
Wolf-Hirschhorn syndrome
PKU due to dihydropteridine reductase deficiency
Dentinogenesis imperfecta-1
C3b inactivator deficiency
Aspartylglucosaminuria
Williams-Beuren syndrome, type II
Sclerotylosis
Anterior segment mesenchymal dysgenesis
Pseudohypoaldosteronism
Hepatocellular carcinoma
Factor XI deficiency
Fletcher factor deficiency

(b) Genes on human chromosome 4 that are associated with disease when mutated

FIGURE 1.1 **The Human Genome Project. (a)** The human genome is a complete set of human chromosomes. People have two sets of chromosomes, one from each parent. Collectively, each set of chromosomes is composed of a DNA sequence that is approximately 3 billion nucleotide base pairs long. Estimates suggest that each set contains about 20,000 to 25,000 different genes. This figure emphasizes the DNA found in the cell nucleus. Humans also have a small amount of DNA in their mitochondria, which has also been sequenced. **(b)** An important outcome of this genetic research is the identification of genes that contribute to human diseases. This illustration depicts a map of a few genes that are located on human chromosome 4. When these genes carry certain rare mutations, they can cause the diseases designated in this figure.

Concept check: How might a better understanding of our genes be used in the field of medicine?

New genetic technologies are often met with skepticism and sometimes even with disdain. An example is DNA fingerprinting, a molecular method for identifying an individual based on a DNA sample (see Chapter 25). Though this technology is now relatively common in the area of forensic science, it was not always universally accepted. High-profile crime cases in the news cause us to realize that not everyone accepts the accuracy of DNA fingerprinting, in spite of its extraordinary ability to uniquely identify individuals.

A second controversial example is mammalian cloning. In 1997, Ian Wilmut and his colleagues produced clones of sheep, using mammary cells from an adult animal (**Figure 1.2**). More recently, such cloning has been achieved in several mammalian species, including cows, mice, goats, pigs, and cats. In 2002, the first pet was cloned, a cat named Carbon copy, or Copycat (see photo at the beginning of the chapter). The cloning of mammals provides the potential for many practical applications. Cloning of livestock would enable farmers to use cells from their best individuals to create genetically homogeneous herds. This could be advantageous in terms of agricultural yield, although such a genetically homogeneous herd may be more susceptible to certain diseases. However, people have become greatly concerned with the possibility of human cloning. As discussed in Chapter 20, this prospect has raised serious ethical questions. Within the past few years, legislative bills have been introduced that involve bans on human cloning.

Finally, genetic technologies provide the means of modifying the traits of animals and plants in ways that would have been unimaginable just a few decades ago. **Figure 1.3a** illustrates a bizarre example in which scientists introduced a gene from jellyfish into mice. Certain species of jellyfish emit a "green glow" produced by a gene that encodes a bioluminescent protein called green fluorescent protein (GFP). When exposed to blue or ultraviolet (UV) light, the protein emits a striking green-colored light. Scientists were able to clone the *GFP* gene from a sample of jellyfish cells and then introduce this gene into laboratory

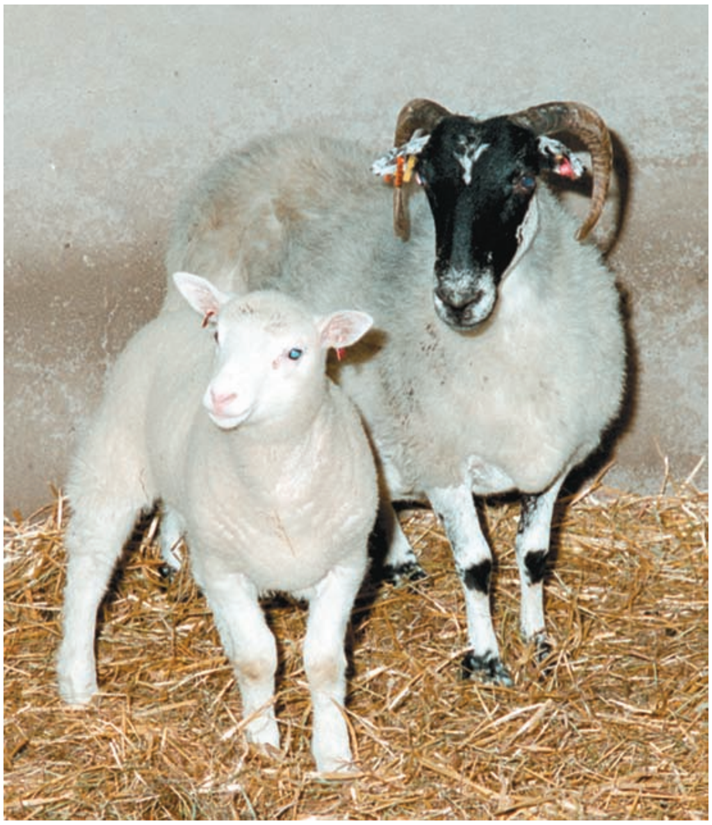
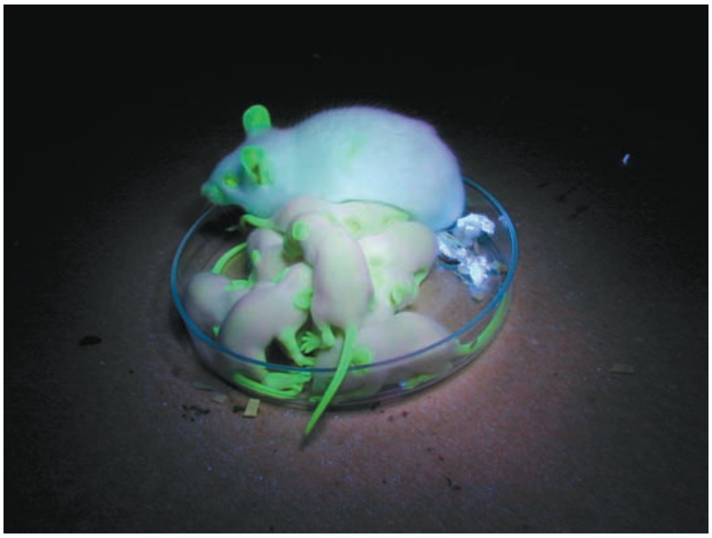
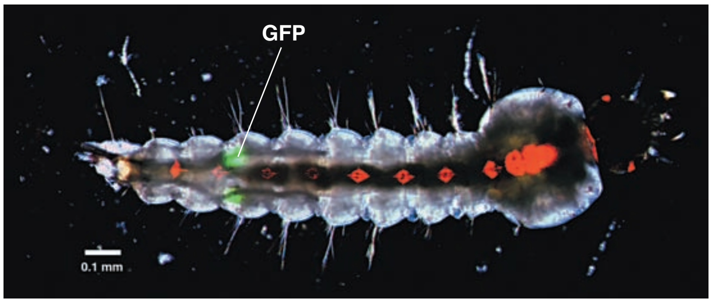


FIGURE 1.2 **The cloning of a mammal.** The lamb on the left is Dolly, the first mammal to be cloned. She was cloned from a cell of a Finn Dorset (a white-faced sheep). The sheep on the right is Dolly's surrogate mother, a Blackface ewe. A description of how Dolly was produced is presented in Chapter 20.

Concept check: What ethical issues may be associated with human cloning?



(a) GFP expressed in mice



(b) GFP expressed in the gonads of a male mosquito

FIGURE 1.3 **The introduction of a jellyfish gene into laboratory mice and mosquitoes.** **(a)** A gene that naturally occurs in the jellyfish encodes a protein called green fluorescent protein (GFP). The GFP gene was cloned and introduced into mice. When these mice are exposed to ultraviolet light, GFP emits a bright green color. These mice glow green, just like jellyfish! **(b)** GFP was introduced next to a gene sequence that causes the expression of GFP only in the gonads of male mosquitoes. This allows researchers to identify and sort males from females.

Concept check: Why is it useful to sort male from female mosquitoes?

mice. The green fluorescent protein is made throughout the cells of their bodies. As a result, their skin, eyes, and organs give off an eerie green glow when exposed to UV light. Only their fur does not glow.

The expression of green fluorescent protein allows researchers to identify particular proteins in cells or specific body parts. For example, Andrea Crisanti and colleagues have altered mosquitoes to express GFP only in the gonads of males (**Figure 1.3b**). This enables the researchers to identify and sort males from females. Why is this useful? The ability to rapidly sort mosquitoes by sex makes it possible to produce populations of sterile males and then release the sterile males without the risk of releasing additional females. The release of sterile males may be an effective means of controlling mosquito populations because females breed only once before they die. Mating with a sterile male prevents a female from producing offspring. In 2008, Osamu Shimomura, Martin Chalfie, and Roger Tsien received the Nobel Prize in chemistry for the discovery and the development of GFP, which has become a widely used tool in biology.

Overall, as we move forward in the twenty-first century, the excitement level in the field of genetics is high, perhaps higher than it has ever been. Nevertheless, the excitement generated by new genetic knowledge and technologies will also create many ethical and societal challenges. In this chapter, we begin with an overview of genetics and then explore the various fields of genetics and their experimental approaches.

1.1 THE MOLECULAR EXPRESSION OF GENES

Genetics is the branch of biology that deals with heredity and variation. It stands as the unifying discipline in biology by allowing us to understand how life can exist at all levels of complexity, ranging from the molecular to the population level. Genetic variation is the root of the natural diversity that we observe among members of the same species as well as among different species.

Genetics is centered on the study of genes. A gene is classically defined as a unit of heredity, but such a vague definition does not do justice to the exciting characteristics of genes as intricate molecular units that manifest themselves as critical contributors to cell structure and function. At the molecular level, a **gene** is a segment of DNA that has the information to produce a functional product. The functional product of most genes is a polypeptide—a linear sequence of amino acids that folds into units that constitute proteins. In addition, genes are commonly described according to the way they affect **traits,** which are the characteristics of an organism. In humans, for example, we speak of traits such as eye color, hair texture, and height. An ongoing theme of this textbook is the relationship between genes and traits. As an organism grows and develops, its collection of genes provides a blueprint that determines its characteristics.

In this section, we examine the general features of life with an emphasis on the molecular level. As will become apparent, genetics is the common thread that explains the existence of life and its continuity from generation to generation. For most students, this chapter should serve as a cohesive review of topics they learned in other introductory courses such as general biology. Even so, it is usually helpful to see the "big picture" of genetics before delving into the finer details that are covered in Chapters 2 through 27.

Living Cells Are Composed of Biochemicals

To fully understand the relationship between genes and traits, we need to begin with an examination of the composition of living organisms. Every cell is constructed from intricately organized chemical substances. Small organic molecules such as glucose and amino acids are produced from the linkage of atoms via chemical bonds. The chemical properties of organic molecules are essential for cell vitality in two key ways. First, the breaking of chemical bonds during the degradation of small molecules provides energy to drive cellular processes. A second important function of these small organic molecules is their role as the building blocks for the synthesis of larger molecules. Four important categories of larger cellular molecules are **nucleic acids** (i.e., DNA and RNA), **proteins, carbohydrates,** and **lipids.** Three of these—nucleic acids, proteins, and carbohydrates—form **macromolecules** that are composed of many repeating units of smaller building blocks. Proteins, RNA, and carbohydrates can be made from hundreds or even thousands of repeating building blocks. DNA is the largest macromolecule found in living cells. A single DNA molecule can be composed of a linear sequence of hundreds of millions of nucleotides!

The formation of cellular structures relies on the interactions of molecules and macromolecules. For example, nucleotides are the building blocks of DNA, which is one component of chromosomes (**Figure 1.4**). Besides DNA, different types of proteins are important for the proper structure of chromosomes. Within a eukaryotic cell, the chromosomes are contained in a compartment called the cell nucleus. The nucleus is bounded by a double membrane composed of lipids and proteins that shields the chromosomes from the rest of the cell. The organization of chromosomes within a cell nucleus protects the chromosomes from mechanical damage and provides a single compartment for genetic activities such as gene transcription. As a general theme, the formation of large cellular structures arises from interactions among different molecules and macromolecules. These cellular structures, in turn, are organized to make a complete living cell.

Each Cell Contains Many Different Proteins That Determine Cell Structure and Function

To a great extent, the characteristics of a cell depend on the types of proteins that it makes. All of the proteins that a cell or organism makes at a given time is called its **proteome.** As we will learn throughout this textbook, proteins are the "workhorses" of all living cells. The range of functions among different types of proteins is truly remarkable. Some proteins help determine the shape and structure of a given cell. For example, the protein known as tubulin can assemble into large structures known as microtubules, which provide the cell with internal structure and organization. Other proteins are inserted into cell membranes and aid in the transport of ions and small molecules across the

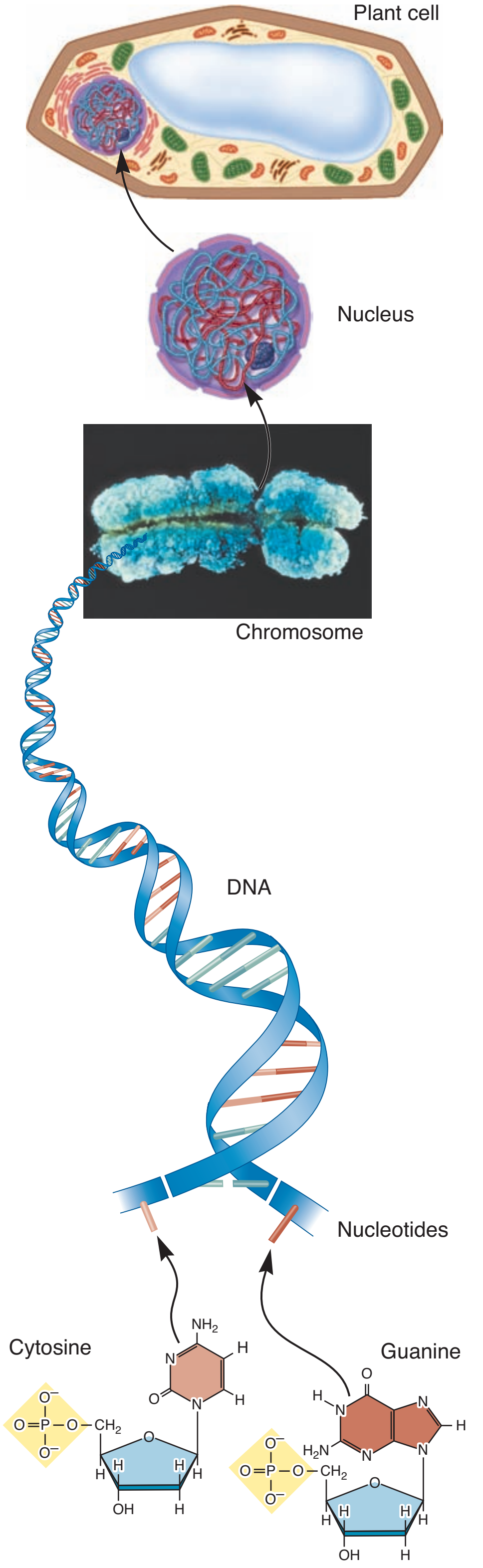


FIGURE 1.4 **Molecular organization of a living cell.** Cellular structures are constructed from smaller building blocks. In this example, DNA is formed from the linkage of nucleotides to produce a very long macromolecule. The DNA associates with proteins to form a chromosome. The chromosomes are located within a membrane-bound organelle called the nucleus, which, along with many different types of organelles, is found within a complete cell.

Concept check: Is DNA a small molecule, a macromolecule, or an organelle?

membrane. Proteins may also function as biological motors. An interesting case is the protein known as myosin, which is involved in the contractile properties of muscle cells. Within multicellular organisms, certain proteins also function in cell-to-cell recognition and signaling. For example, hormones such as insulin are secreted by endocrine cells and bind to the insulin receptor protein found within the plasma membrane of target cells.

Enzymes, which accelerate chemical reactions, are a particularly important category of proteins. Some enzymes play a role in the breakdown of molecules or macromolecules into smaller units. These are known as catabolic enzymes and are important in the utilization of energy. Alternatively, anabolic enzymes and accessory proteins function in the synthesis of molecules and macromolecules throughout the cell. The construction of a cell greatly depends on its proteins involved in anabolism because these are required to synthesize all cellular macromolecules.

Molecular biologists have come to realize that the functions of proteins underlie the cellular characteristics of every organism. At the molecular level, proteins can be viewed as the active participants in the enterprise of life.

DNA Stores the Information for Protein Synthesis

As mentioned, the genetic material of living organisms is composed of a substance called deoxyribonucleic acid, abbreviated DNA. The DNA stores the information needed for the synthesis of all cellular proteins. In other words, the main function of the genetic blueprint is to code for the production of proteins in the correct cell, at the proper time, and in suitable amounts. This is an extremely complicated task because living cells make thousands of different proteins. Genetic analyses have shown that a typical bacterium can make a few thousand different proteins, and estimates among higher eukaryotes range in the tens of thousands.

DNA's ability to store information is based on its structure. DNA is composed of a linear sequence of **nucleotides,** each of which contains one of four nitrogen-containing bases: adenine (A), thymine (T), guanine (G), or cytosine (C). The linear order of these bases along a DNA molecule contains information similar to the way that groups of letters of the alphabet represent words. For example, the "meaning" of the sequence of bases ATGGGCCTTAGC differs from that of TTTAAGCTTGCC. DNA sequences within most genes contain the information to direct the order of amino acids within **polypeptides** according to the **genetic code.** In the code, a three-base sequence specifies one particular **amino acid** among the 20 possible choices. One or more polypeptides form a functional protein. In this way, the DNA can store the information to specify the proteins made by an organism.

DNA Sequence	Amino Acid Sequence
ATG GGC CTT AGC	Methionine Glycine Leucine Serine
TTT AAG CTT GCC	Phenylalanine Lysine Leucine Alanine

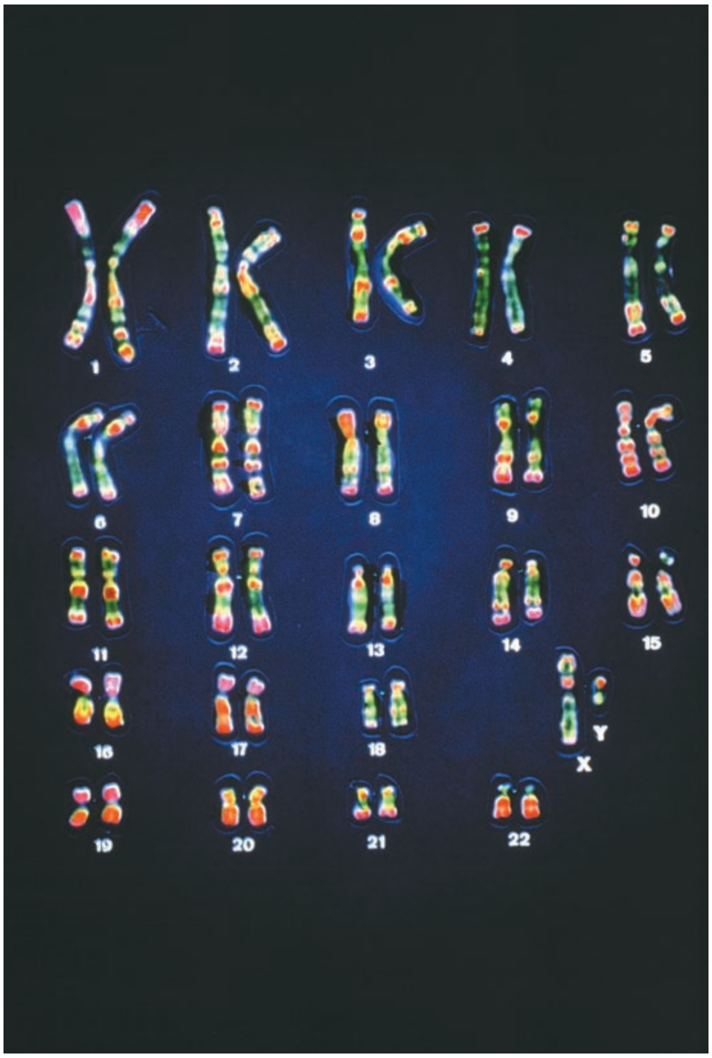


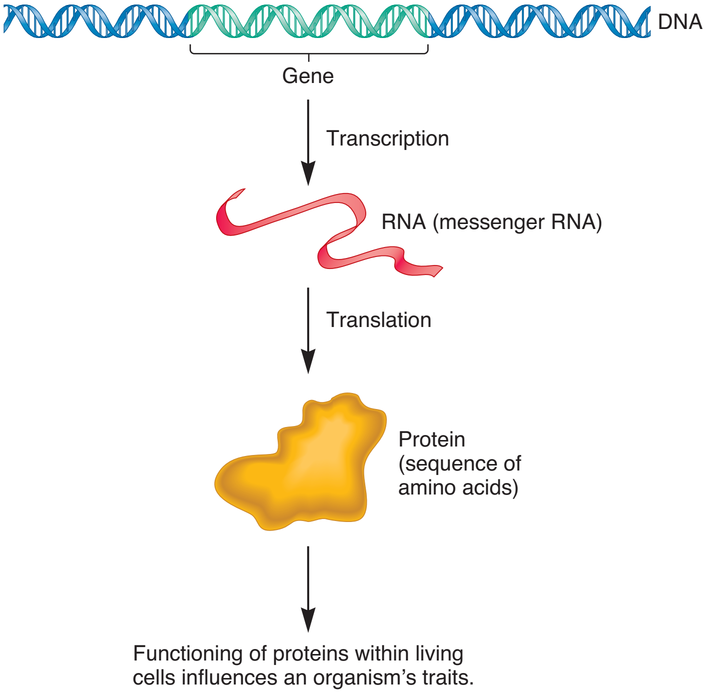
FIGURE 1.5 **A micrograph of the 46 chromosomes found in a cell from a human male.**

Concept check: Which types of macromolecules are found in chromosomes?

In living cells, the DNA is found within large structures known as **chromosomes.** Figure 1.5 is a micrograph of the 46 chromosomes contained in a cell from a human male, which are found in pairs. The DNA of an average human chromosome is an extraordinarily long, linear, double-stranded structure that contains well over a hundred million nucleotides. Along the immense length of a chromosome, the genetic information is parceled into functional units known as genes. An average-sized human chromosome is expected to contain about 1000 different genes.

The Information in DNA Is Accessed During the Process of Gene Expression

To synthesize its proteins, a cell must be able to access the information that is stored within its DNA. The process of using a gene sequence to affect the characteristics of cells and organisms is referred to as **gene expression.** At the molecular level, the information within genes is accessed in a stepwise process. In the first step, known as **transcription,** the DNA sequence within a gene is copied into a nucleotide sequence of **ribonucleic acid (RNA).** Most genes encode RNAs that contain the information for the synthesis of a particular polypeptide. This type of RNA is called **messenger RNA (mRNA).** During the process of **translation,** the sequence of nucleotides in an mRNA provides the information (using the genetic code) to produce the amino acid sequence of a polypeptide (Figure 1.6). After a polypeptide is made, it folds into a three-dimensional structure. As mentioned, a protein is a functional unit. Some proteins are composed of a single polypeptide, and other proteins consist of two or more polypeptides. Some RNA molecules are not mRNA molecules and therefore are not translated into polypeptides. We will consider the functions of these RNA molecules in Chapter 15 (see Table 15.1).



ONLINE ANIMATION

FIGURE 1.6 **Gene expression at the molecular level.** The expression of a gene is a multistep process. During transcription, one of the DNA strands is used as a template to make an RNA strand. During translation, the RNA strand is used to specify the sequence of amino acids within a polypeptide. One or more polypeptides produce a protein that functions within the cell, thereby influencing an organism's traits.

Concept check: Where is the information to make a polypeptide stored?

The expression of most genes results in the production of proteins with specific structures and functions. The unique relationship between gene sequences and protein structures is of paramount importance because the distinctive structure of each protein determines its function within a living cell or organism. Mediated by the process of gene expression, therefore, the sequence of nucleotides in DNA stores the information required for synthesizing proteins with specific structures and functions.

1.1 REVIEWING THE KEY CONCEPTS

- Living cells are composed of nucleic acids (DNA and RNA), proteins, carbohydrates, and lipids. The proteome largely determines the structure and function of cells (see Figure 1.4).
- DNA, which is found within chromosomes, stores the information to make proteins (see Figure 1.5).
- Most genes encode polypeptides that are units within functional proteins. Gene expression at the molecular level involves transcription to produce mRNA and translation to produce a polypeptide (see Figure 1.6).

1.1 COMPREHENSION QUESTIONS

1. Which of the following is not a constituent of a cell's proteome?
   a. An enzyme
   b. A motor protein
   c. A receptor in the plasma membrane
   d. An mRNA
2. A gene is a segment of DNA that has the information to produce a functional product. The functional product of most genes is
   a. DNA.
   b. mRNA.
   c. a polypeptide.
   d. All of the above.
3. The function of the genetic code is to
   a. promote transcription.
   b. specify the amino acids within a polypeptide.
   c. alter the sequence of DNA.
   d. None of the above.
4. The process of transcription directly results in the synthesis of
   a. DNA.
   b. RNA.
   c. a polypeptide.
   d. All of the above.

1.2 THE RELATIONSHIP BETWEEN GENES AND TRAITS

A trait is any characteristic that an organism displays. In genetics, we often focus our attention on **morphological traits** that affect the appearance, form, and structure of an organism. The color of a flower and the height of a pea plant are morphological traits. Geneticists frequently study these types of traits because they are easy to evaluate. For example, an experimenter can simply look at a plant and tell if it has red or white flowers. However, not all traits are morphological. **Physiological traits** affect the ability of an organism to function. For example, the rate at which a bacterium metabolizes a sugar such as lactose is a physiological trait. Like morphological traits, physiological traits are controlled, in part, by the expression of genes. **Behavioral traits** also affect the ways an organism responds to its environment. An example is the mating calls of bird species. In animals, the nervous system plays a key role in governing such traits. In this section, we will examine the relationship between the expression of genes and an organism's traits.

The Molecular Expression of Genes Within Cells Leads to an Organism's Traits

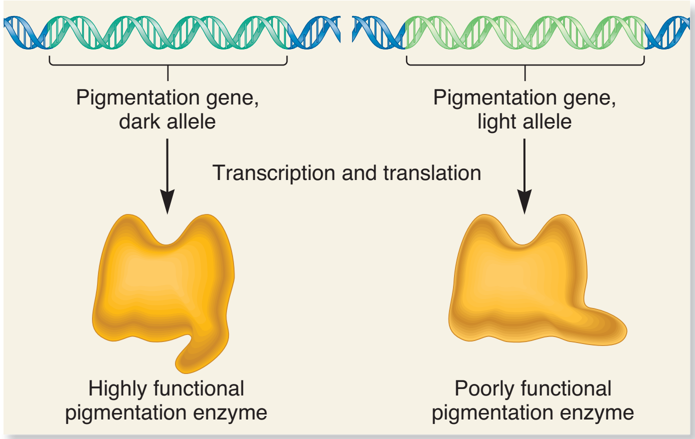
A complicated, yet very exciting, aspect of genetics is that our observations and theories span four levels of biological organization: molecules, cells, organisms, and populations. This can make it difficult to appreciate the relationship between genes and traits. To understand this connection, we need to relate the following phenomena:

1. As we learned in Section 1.1, genes are expressed at the **molecular level.** In other words, gene transcription and translation lead to the production of a particular protein, which is a molecular process.
2. Proteins often function at the **cellular level.** The function of a protein within a cell affects the structure and workings of that cell.
3. An organism's traits are determined by the characteristics of its cells. We do not have microscopic vision, yet when we view morphological traits, we are really observing the properties of an individual's cells. For example, a red flower has its color because the flower cells make a red pigment. The trait of red flower color is an observation at the **organism level,** yet the trait is rooted in the molecular characteristics of the organism's cells.
4. A **species** is a group of organisms that maintains a distinctive set of attributes in nature. The occurrence of a trait within a species is an observation at the **population level.** Along with learning how a trait occurs, we also want to understand why a trait becomes prevalent in a particular species. In many cases, researchers discover that a trait predominates within a population because it promotes the reproductive success of the members of the population. This leads to the evolution of beneficial traits.

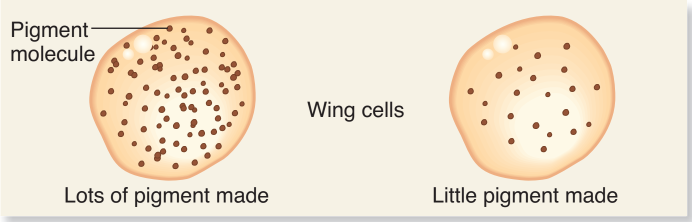
As a schematic example to illustrate the four levels of genetics, **Figure 1.7** shows the trait of pigmentation in butterflies. One is light-colored and the other is very dark. Let's consider how we can explain this trait at the molecular, cellular, organism, and population levels.

At the molecular level, we need to understand the nature of the gene or genes that govern this trait. As shown in Figure 1.7a, a gene, which we will call the pigmentation gene, is responsible for the amount of pigment produced. The pigmentation gene can exist in two different forms called **alleles.** In this example, one allele confers a dark pigmentation and one causes a light pigmentation. Each of these alleles encodes a protein that functions as a pigment-synthesizing enzyme. However, the DNA sequences of the two alleles differ slightly from each other. This difference in the DNA sequence leads to a variation in the structure and function of the respective pigmentation enzymes.

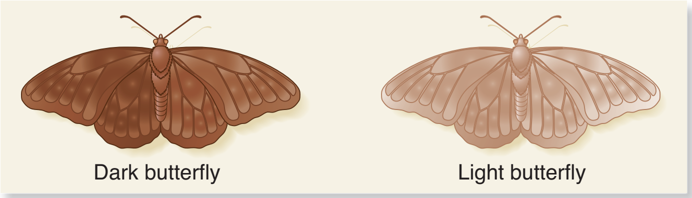
At the cellular level (Figure 1.7b), the functional differences between the pigmentation enzymes affect the amount of pigment produced. The allele causing dark pigmentation, which


Pigmentation gene,
dark allele
Pigmentation gene,
light allele
Transcription and translation
Highly functional
pigmentation enzyme
Poorly functional
pigmentation enzyme

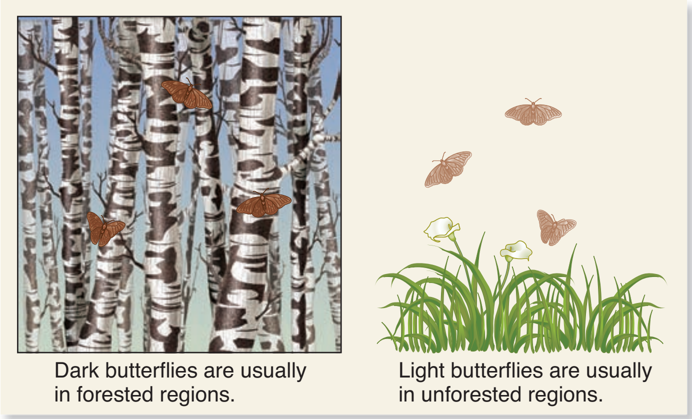
(a) Molecular level


Pigment molecule
Wing cells
Lots of pigment made
Little pigment made

(b) Cellular level


Dark butterfly
Light butterfly

(c) Organism level


Dark butterflies are usually
in forested regions.
Light butterflies are usually
in unforested regions.

(d) Population level

FIGURE 1.7 **The relationship between genes and traits at the (a) molecular, (b) cellular, (c) organism, and (d) population levels.**

Concept check: Which butterfly has a more active pigment-producing enzyme, the light- or dark-colored one?

is shown on the left, encodes a protein that functions very well. Therefore, when this gene is expressed in the cells of the wings, a large amount of pigment is made. By comparison, the allele causing light pigmentation encodes an enzyme that functions poorly. Therefore, when this allele is the only pigmentation gene expressed, little pigment is made.

At the organism level (Figure 1.7c), the amount of pigment in the wing cells governs the color of the wings. If the pigment cells produce high amounts of pigment, the wings are dark-colored; if the pigment cells produce little pigment, the wings are light.

Finally, at the population level (Figure 1.7d), geneticists want to know why a species of butterfly has some members with dark wings and others with light wings. One possible explanation is differential predation. The butterflies with dark wings might avoid being eaten by birds if they happen to live within the dim light of a forest. The dark wings help to camouflage the butterfly if it were perched on a dark surface such as a tree trunk. In contrast, the lightly colored wings would be an advantage if the butterfly inhabited a brightly lit meadow. Under these conditions, a bird may be less likely to notice a light-colored butterfly that is perched on a sunlit surface. A geneticist might study this species of butterfly and find that the dark-colored members usually live in forested areas and the light-colored members reside in unforested regions.

Inherited Differences in Traits Are Due to Genetic Variation

In Figure 1.7, we considered how gene expression could lead to variation in a trait of an organism, such as dark- versus light-colored butterflies. Variation in traits among members of the same species is very common. For example, some people have brown hair, and others have blond hair; some petunias have white flowers, but others have purple flowers. These are examples of **genetic variation.** This term describes the differences in inherited traits among individuals within a population.

In large populations that occupy a wide geographic range, genetic variation can be quite striking. In fact, morphological differences have often led geneticists to misidentify two members of the same species as belonging to separate species. As an example, **Figure 1.8** shows two dyeing poison frogs that are members of the same species, *Dendrobates tinctorius*. They display dramatic differences in their markings. Such contrasting forms within a single species are termed **morphs.** You can easily imagine how someone might mistakenly conclude that these frogs are not members of the same species.

Changes in the nucleotide sequence of DNA underlie the genetic variation that we see among individuals. Throughout this textbook, we will routinely examine how variation in the genetic material results in changes in the outcome of traits. At the molecular level, genetic variation can be attributed to different types of modifications.

1. Small or large differences can occur within gene sequences. When such changes initially occur, they are called **gene mutations,** which are heritable changes in

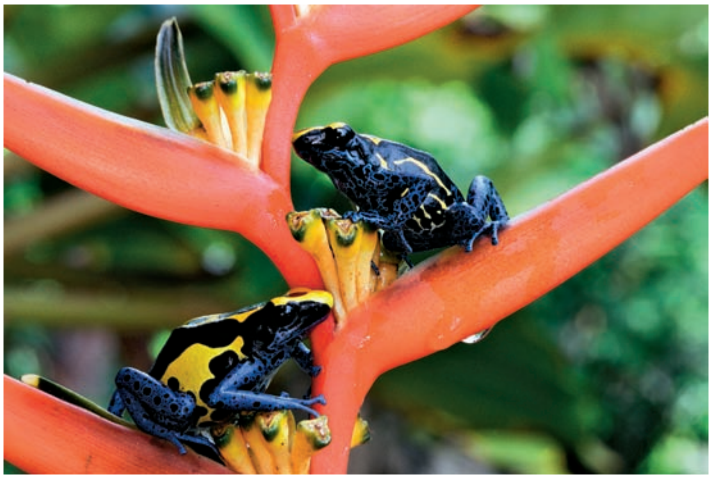


FIGURE 1.8 **Two dyeing poison frogs (*Dendrobates tinctorius*) showing different morphs within a single species.**

Concept check: Why do these two frogs look so different?

the genetic material. Gene mutations result in genetic variation in which a gene is found in two or more alleles, as previously described in Figure 1.7. In many cases, gene mutations alter the expression or function of the protein that the gene specifies.

2. Major alterations can also occur in the structure of a chromosome. A large segment of a chromosome can be lost, rearranged, or reattached to another chromosome.
3. Variation may also occur in the total number of chromosomes. In some cases, an organism may inherit one too many or one too few chromosomes. In other cases, it may inherit an extra set of chromosomes.

Variations within the sequences of genes are a common source of genetic variation among members of the same species. In humans, familiar examples of variation involve genes for eye color, hair texture, and skin pigmentation. Chromosome variation—a change in chromosome structure or number (or both)—is also found, but this type of change is often detrimental. Many human genetic disorders are the result of chromosomal alterations. An example is Down syndrome, which is due to the presence of an extra chromosome (**Figure 1.9a**). By comparison, chromosome variation in plants is common and often can lead to plants with superior characteristics, such as increased resistance to disease. Plant breeders have frequently exploited this observation. Cultivated varieties of wheat, for example, have many more chromosomes than the wild species (**Figure 1.9b**).

Traits Are Governed by Genes and by the Environment

In our discussion thus far, we have considered the role that genes play in the outcome of traits. Another critical factor is the **environment**—the surroundings in which an organism exists. A variety of factors in an organism's environment profoundly affect its morphological and physiological features. For example, a person's diet greatly influences many traits such as height, weight, and even intelligence. Likewise, the amount of sunlight a plant receives affects its growth rate and the color of its flowers. The term **norm of reaction** refers to the effects of environmental variation on an individual's traits.

External influences may dictate the way that genetic variation is manifested in an individual. An interesting example is the human genetic disease **phenylketonuria (PKU).** Humans possess a gene that encodes an enzyme known as phenylalanine hydroxylase. Most people have two functional copies of this gene. People with one or two functional copies of the gene can eat foods containing the amino acid phenylalanine and metabolize it properly.

A rare variation in the sequence of the phenylalanine hydroxylase gene results in a nonfunctional version of this protein. Individuals with two copies of this rare, inactive allele cannot metabolize phenylalanine properly. This occurs in about 1 in 8000 births among Caucasians in the United States. When

1923
1989 Summer Games
(a)
(b)

FIGURE 1.9 **Examples of chromosome variation.** **(a)** A person with Down syndrome competing in the Special Olympics. This person has 47 chromosomes rather than the common number of 46, because she has an extra copy of chromosome 21. **(b)** A wheat plant. Bread wheat is derived from the contributions of three related species with two sets of chromosomes each, producing an organism with six sets of chromosomes.

Concept check: Are these examples of gene mutations, variation in chromosome structure, or variation in chromosome number?

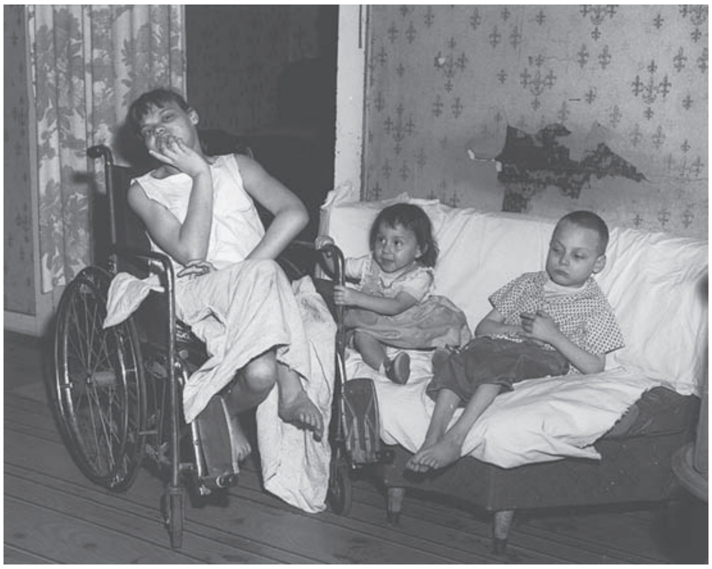


FIGURE 1.10 **Environmental influence on the outcome of PKU within a single family.** All three children pictured here have inherited the alleles that cause PKU. The child in the middle was raised on a phenylalanine-free diet and developed normally. The other two children were born before the benefits of a phenylalanine-free diet were known and were raised on diets that contained phenylalanine. Therefore, they manifest a variety of symptoms, including mental impairment. People born today with this disorder are usually diagnosed when infants. (*Photo from the March of Dimes Birth Defects Foundation.*)

Concept check: What was different about the environment of the young girl in the center versus her sister and brother? How did this affect her traits?

given a standard diet containing phenylalanine, individuals with this disorder are unable to break down this amino acid. Phenylalanine accumulates and is converted into phenylketones, which are detected in the urine. PKU individuals manifest a variety of detrimental traits, including mental impairment, underdeveloped teeth, and foul-smelling urine. In contrast, when PKU individuals are identified at birth and raised on a restricted diet that is low in phenylalanine, they develop normally (**Figure 1.10**). Fortunately, through routine newborn screening, most affected babies in the United States are now diagnosed and treated early. PKU provides a dramatic example of how the environment and an individual's genes can interact to influence the traits of the organism.

During Reproduction, Genes Are Passed from Parent to Offspring

Now that we have considered how genes and the environment govern the outcome of traits, we can turn to the issue of inheritance. How are traits passed from parents to offspring? The foundation for our understanding of inheritance came from the studies of pea plants by Gregor Mendel in the nineteenth century. His work revealed that genetic determinants, which we now call genes, are passed from parent to offspring as discrete units. We can predict the outcome of many genetic crosses based on Mendel's laws of inheritance.

The inheritance patterns identified by Mendel can be explained by the existence of chromosomes and their behavior during cell division. As in Mendel's pea plants, sexually reproducing species are commonly **diploid.** This means they contain two copies of each chromosome, one from each parent. The two copies are called **homologs** of each other. Because genes are located within chromosomes, diploid organisms have two copies of most genes. Humans, for example, have 46 chromosomes, which are found in homologous pairs (**Figure 1.11a**). With the exception of the sex chromosomes (X and Y), each homologous pair contains the same kinds of genes. For example, both copies of human chromosome 12 carry the gene that encodes phenylalanine hydroxylase, which was discussed previously. Therefore, an individual has two copies of this gene that may or may not be identical alleles.

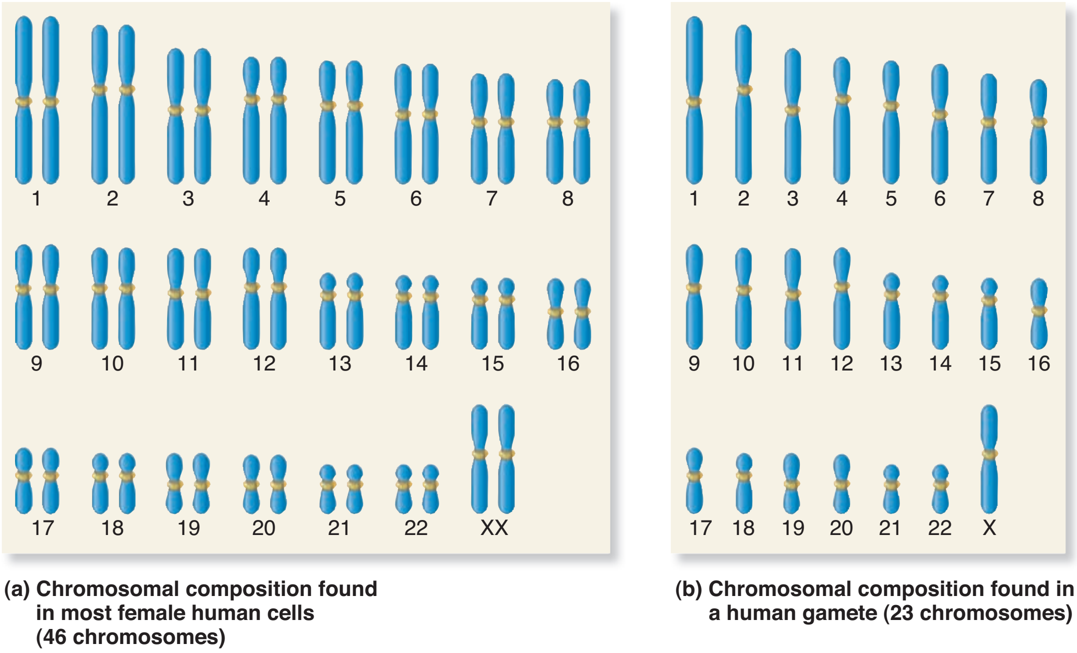
Most cells of the human body that are not directly involved in sexual reproduction contain 46 chromosomes. These cells are called **somatic cells.** In contrast, the **gametes**—sperm and egg cells—contain half that number (23) and are termed **haploid** (**Figure 1.11b**). The union of gametes during fertilization restores the diploid number of chromosomes. The primary advantage of sexual reproduction is that it enhances genetic variation. For example, a tall person with blue eyes and a short person with brown eyes may have short offspring with blue eyes or tall offspring with brown eyes. Therefore, sexual reproduction can result in new combinations of two or more traits that differ from those of either parent.

The Genetic Composition of a Species Evolves over the Course of Many Generations

As we have just seen, sexual reproduction has the potential to enhance genetic variation. This can be an advantage for a population of individuals as they struggle to survive and compete within their natural environment. The term **biological evolution,** or simply, **evolution,** refers to the phenomenon that the genetic makeup of a population can change from one generation to the next.

As suggested by Charles Darwin, the members of a species are in competition with one another for essential resources. Random genetic changes (i.e., mutations) occasionally occur within an individual's genes, and sometimes these changes lead to a modification of traits that promote reproductive success. For example, over the course of many generations, random gene mutations have lengthened the neck of the giraffe, enabling it to feed on leaves located higher in trees. When a mutation creates a new allele that is beneficial, the allele may become prevalent in future generations because the individuals carrying the allele are more likely to survive and reproduce and pass the beneficial allele to their offspring. This process is known as **natural selection.** In this way, a species becomes better adapted to survive and reproduce in its native environment.

Over a long period of time, the accumulation of many genetic changes may lead to rather striking modifications in a



(a) Chromosomal composition found in most female human cells (46 chromosomes)

(b) Chromosomal composition found in a human gamete (23 chromosomes)

FIGURE 1.11 **The complement of human chromosomes in somatic cells and gametes. (a)** A schematic drawing of the 46 chromosomes of a human. With the exception of the sex chromosomes, these are always found in homologous pairs. **(b)** The chromosomal composition of a gamete, which contains only 23 chromosomes, one from each pair. This gamete contains an X chromosome. Half of the gametes from human males contain a Y chromosome instead of the X chromosome.

Concept check: The leaf cells of a corn plant contain 20 chromosomes each. How many chromosomes are found in a gamete made by a corn plant?

species' characteristics. As an example, **Figure 1.12** depicts the evolution of the modern-day horse. A variety of morphological changes occurred, including an increase in size, fewer toes, and modified jaw structure. The changes can be attributed to natural selection producing adaptations to changing global climates. Over North America, where much of horse evolution occurred, large areas of dense forests were replaced with grasslands. The increase in size and changes in foot structure enabled horses to escape predators more easily and travel greater distances in search of food. The changes seen in horses' teeth are consistent with a shift from eating tender leaves to eating grasses and other vegetation that are more abrasive and require more chewing.

1.2 REVIEWING THE KEY CONCEPTS

- Genetics, which governs an organism's traits, spans the molecular, cellular, organism, and population levels (see Figure 1.7).
- Genetic variation underlies variation in traits. In addition, the environment plays a key role (see Figures 1.8–1.10).
- During reproduction, genetic material is passed from parents to offspring. In many species, somatic cells are diploid and have two sets of chromosomes, whereas gametes are haploid and have a single set (see Figure 1.11).
- Evolution refers to a change in the genetic composition of a population from one generation to the next (see Figure 1.12).

1.2 COMPREHENSION QUESTIONS

1. Gene expression can be viewed at which of the following levels?
   a. Molecular and cellular levels
   b. Organism level
   c. Population level
   d. All of the above
2. Variation in the traits of organisms may be attributable to
   a. gene mutations.
   b. alterations in chromosome structure.
   c. variation in chromosome number.
   d. All of the above.
3. A human skin cell has 46 chromosomes. A human sperm cell has
   a. 23.
   b. 46.
   c. 92.
   d. None of the above.
4. Evolutionary change caused by natural selection results in species with
   a. greater complexity.
   b. less complexity.
   c. greater reproductive success in their native environment.
   d. the ability to survive longer.

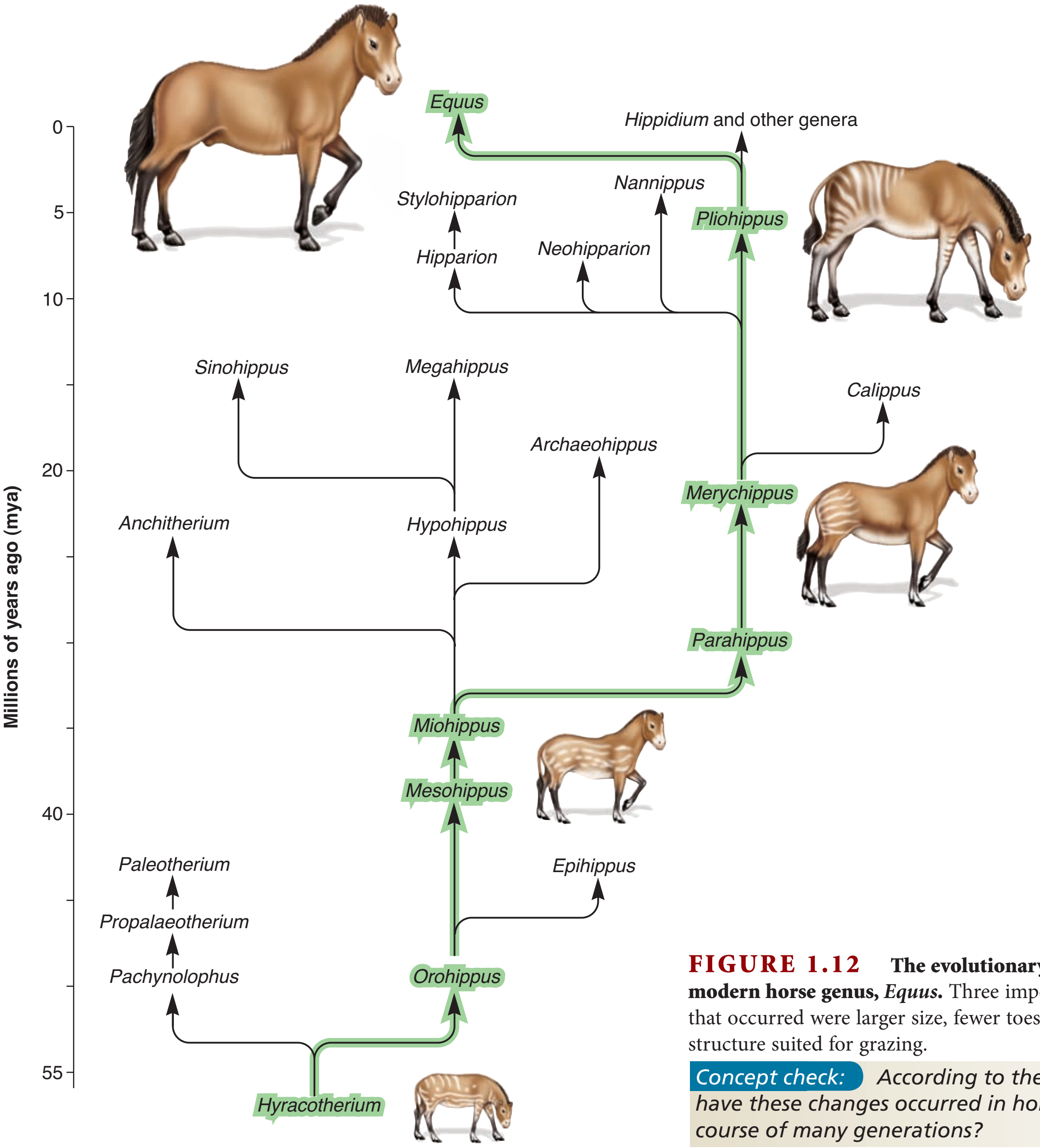

Millions of years ago (mya)
0
5
10
20
40
55
Equus
Hippidium and other genera
Stylohipparion
Nannippus
Pliohippus
Hipparion
Neohipparion
Sinohippus
Megahippus
Calippus
Archaeohippus
Merychippus
Anchitherium
Hypohippus
Parahippus
Miohippus
Mesohippus
Paleotherium
Epihippus
Propalaeotherium
Pachynolophus
Orohippus
Hyracotherium

FIGURE 1.12 **The evolutionary changes that led to the modern horse genus, *Equus*.** Three important morphological changes that occurred were larger size, fewer toes, and a shift toward a jaw structure suited for grazing.

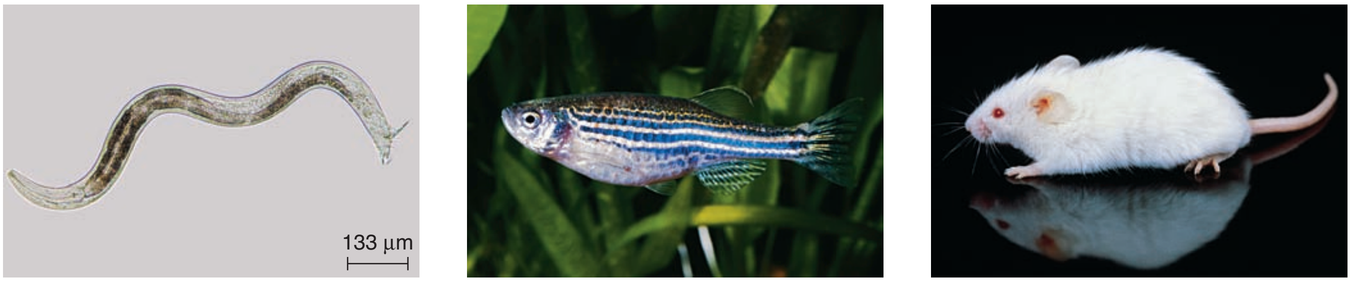
Concept check: According to the theory of evolution, why have these changes occurred in horse populations over the course of many generations?

1.3 FIELDS OF GENETICS

Genetics is a broad discipline encompassing molecular, cellular, organism, and population biology. Many scientists who are interested in genetics have been trained in supporting disciplines such as biochemistry, biophysics, cell biology, mathematics, microbiology, population biology, ecology, agriculture, and medicine. Experimentally, geneticists often focus their efforts on **model organisms**—organisms studied by many different researchers so they can compare their results and determine scientific principles that apply more broadly to other species. Figure 1.13 shows some common examples, including *Escherichia coli* (a bacterium), *Saccharomyces cerevisiae* (a yeast), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (a nematode worm), *Danio rerio* (zebrafish), *Mus musculus* (mouse), and *Arabidopsis thaliana* (a flowering plant). Model organisms offer experimental advantages over other species. For example, *E. coli* is a very simple organism that can be easily grown in the laboratory. By limiting their work to a few such model organisms, researchers can more easily unravel the genetic mechanisms that govern the traits of a given species. Furthermore, the genes found in model organisms often function in a similar way to those found in humans.

The study of genetics has been traditionally divided into three areas—transmission, molecular, and population genetics—although overlap is found among these three fields. In this section, we will examine the general questions that scientists in these areas are attempting to answer.

(a) ***Escherichia coli*** **(b)** ***Saccharomyces cerevisiae*** **(c)** ***Drosophila melanogaster***



(d) ***Caenorhabditis elegans*** **(e)** ***Danio rerio*** **(f)** ***Mus musculus***

(g) ***Arabidopsis thaliana***

FIGURE 1.13 **Examples of model organisms studied by geneticists.** **(a)** *Escherichia coli* (a bacterium), **(b)** *Saccharomyces cerevisiae* (a yeast), **(c)** *Drosophila melanogaster* (fruit fly), **(d)** *Caenorhabditis elegans* (a nematode worm), **(e)** *Danio rerio* (zebrafish), **(f)** *Mus musculus* (mouse), and **(g)** *Arabidopsis thaliana* (a flowering plant).

Concept check: *Can you think of another example of a model organism?*

Transmission Genetics Explores the Inheritance Patterns of Traits as They Are Passed from Parents to Offspring

A scientist working in the field of transmission genetics examines the relationship between the transmission of genes from parent to offspring and the outcome of the offspring's traits. For example, how can two brown-eyed parents produce a blue-eyed child? Or why do tall parents tend to produce tall children, but not always? Our modern understanding of transmission genetics began with the studies of Gregor Mendel. His work provided the conceptual framework for transmission genetics. In particular, he originated the idea that genetic determinants, which we now call genes, are passed as discrete units from parents to offspring via sperm and egg cells. Since these pioneering studies of the 1860s, our knowledge of genetic transmission has greatly increased. Many patterns of genetic transmission are more complex than the simple Mendelian patterns that are described in Chapter 3. The additional complexities of transmission genetics are examined in Chapters 4 through 10.

Experimentally, the fundamental approach of a transmission geneticist is the genetic cross. A **genetic cross** involves breeding two selected individuals and the subsequent analysis of their offspring in an attempt to understand how traits are passed from parents to offspring. In the case of experimental organisms, the researcher chooses two parents with particular traits and then categorizes the offspring according to the traits they possess. In many cases, this analysis is quantitative in nature. For example, an experimenter may cross two tall pea plants and obtain 100 offspring that fall into two categories: 75 tall and 25 dwarf. As we will see in Chapter 3, the ratio of tall and dwarf offspring (3:1) provides important information concerning the inheritance pattern of this trait.

Throughout Chapters 2 to 10, we will learn how researchers seek to answer many fundamental questions concerning the passage of genetic material from cell to cell and the passage of traits from parents to offspring. Here are some of these questions:

How are chromosomes transmitted during cell division and gamete formation? ***Chapter 2***

What are the common patterns of inheritance for genes? ***Chapters 3–5***

Are there unusual patterns of inheritance that cannot be explained by the simple transmission of genes located on chromosomes in the cell nucleus? ***Chapter 6***

When two or more genes are located on the same chromosome, how is the pattern of inheritance affected? ***Chapter 7***

How do variations in chromosome structure or chromosome number occur, and how are they transmitted from parents to offspring? ***Chapter 8***

How are genes transmitted by bacterial species? ***Chapter 9***

How do viruses proliferate? ***Chapter 10***

Molecular Genetics Focuses on a Biochemical Understanding of the Hereditary Material

The goal of molecular genetics, as the name of the field implies, is to understand how the genetic material works at the molecular level. In other words, molecular geneticists want to understand the molecular features of DNA and how these features underlie the expression of genes. The experiments of molecular geneticists are usually conducted within the confines of a laboratory. Their efforts frequently progress to a detailed analysis of DNA, RNA, and protein, using a variety of techniques that are described throughout Parts III, IV, and V of this textbook.

Molecular geneticists often study mutant genes that have abnormal function. This is called a **genetic approach** to the study of a research question. In many cases, researchers analyze the effects of gene mutations that eliminate the function of a gene. This type of mutation is called a **loss-of-function mutation,** and the resulting gene is called a **loss-of-function allele.** By studying the effects of such mutations, the role of the functional, nonmutant gene is often revealed. For example, let's suppose that a particular plant species produces purple flowers. If a loss-of-function mutation within a given gene causes a plant of that species to produce white flowers, one would suspect the role of the functional gene involves the production of purple pigmentation.

Studies within molecular genetics interface with other disciplines such as biochemistry, biophysics, and cell biology. In addition, advances within molecular genetics have shed considerable light on the areas of transmission and population genetics. Our quest to understand molecular genetics has spawned a variety of modern molecular technologies and computer-based approaches. Furthermore, discoveries within molecular genetics have had widespread applications in agriculture, medicine, and biotechnology.

The following are some general questions within the field of molecular genetics:

What are the molecular structures of DNA and RNA? ***Chapter 11***

What is the composition and conformation of chromosomes? ***Chapter 12***

How is the genetic material copied? ***Chapter 13***

How are genes expressed at the molecular level? ***Chapters 14, 15***

How is gene expression regulated so it occurs under the appropriate conditions, in the appropriate cell type, and at the correct stage of development? ***Chapters 16, 17, 24***

What is the molecular nature of mutations? How are mutations repaired? ***Chapter 18***

How have genetic technologies advanced our understanding of genetics? ***Chapters 19, 20***

What is the genetic composition and function of whole genomes? ***Chapters 21, 22***

What is the underlying relationship between genes and genetic diseases? ***Chapter 23***

Population Genetics Is Concerned with Genetic Variation and Its Role in Evolution

The foundations of population genetics arose during the first few decades of the twentieth century. Although many scientists of this era did not accept the findings of Mendel or Darwin, the theories of population genetics provided a compelling way to connect the two viewpoints. Mendel's work and that of many succeeding geneticists gave insight into the nature of genes and how they are transmitted from parents to offspring. The theory of evolution by natural selection proposed by Darwin provided a natural explanation for the variation in characteristics observed among the members of a species. To relate these two phenomena, population geneticists have developed mathematical theories to explain the prevalence of certain alleles within populations of individuals. The work of population geneticists helps us understand how processes such as natural selection have resulted in the prevalence of individuals that carry particular alleles.

Population geneticists are particularly interested in genetic variation and how that variation is related to an organism's environment. In this field, the frequencies of alleles within a population are of central importance. The following are some general questions in population genetics:

Why are two or more different alleles of a gene maintained in a population? ***Chapter 25***

What factors alter the prevalence of alleles within a population? ***Chapter 25***

What are the contributions of genetics and environment in the outcome of a trait? ***Chapter 26***

How do genetics and the environment influence quantitative traits, such as size and weight? ***Chapter 26***

What factors have the most influence on the process of evolution? ***Chapter 27***

How does evolution occur at the molecular level? ***Chapter 27***

Genetics Is an Experimental Science

Science is a way of knowing about our natural world. The science of genetics allows us to understand how the expression of our genes produces the traits that we possess. Regardless of what field of genetics they work in, researchers typically follow two general types of scientific approaches: hypothesis testing and discovery-based science. In **hypothesis testing,** also called the **scientific method,** scientists follow a series of steps to reach verifiable conclusions about the world. Although scientists arrive at their theories in different ways, the scientific method provides a way to validate (or invalidate) a particular hypothesis. Alternatively, research may also involve the collection of data without a preconceived hypothesis. For example, researchers might analyze the genes found in cancer cells to identify those genes that have become mutant. In this case, the scientists may not have a hypothesis about which particular genes may be involved. The collection and analysis of data without the need for a preconceived hypothesis is called **discovery-based science** or, simply, discovery science.

In traditional science textbooks, the emphasis often lies on the product of science. Namely, many textbooks are aimed primarily at teaching the student about the observations scientists have made and the hypotheses they have proposed to explain these observations. Along the way, the student is provided with many bits and pieces of experimental techniques and data. Although this textbook provides you with many observations and hypotheses, it attempts to go one step further. Many of the following chapters contain one or two experiments that have been "dissected" into five individual components to help you to understand the entire scientific process. The five steps are as follows:

1. Background information is provided so you can appreciate what previous observations were known prior to conducting the experiment.
2. Most experiments involve hypothesis testing. In those cases, the figure states the hypothesis the scientists were trying to test. In other words, what scientific question was the researcher trying to answer?
3. Next, the figure follows the experimental steps the scientist took to test the hypothesis. The steps necessary to carry out the experiment are listed in the order in which they were conducted. The figure contains two parallel illustrations labeled *Experimental Level* and *Conceptual Level*. The illustration shown in the Experimental Level helps you to understand the techniques followed. The Conceptual Level helps you to understand what is actually happening at each step in the procedure.
4. The raw data for each experiment are then presented.
5. Last, an interpretation of the data is offered within the text.

The rationale behind this approach is that it will enable you to see the experimental process from beginning to end. Hopefully, you will find this a more interesting and rewarding way to learn about genetics. As you read through the chapters, the experiments will help you to see the relationship between science and scientific theories.

As a student of genetics, you will be given the opportunity to involve your mind in the experimental process. As you are reading an experiment, you may find yourself thinking about alternative approaches and hypotheses. Different people can view the same data and arrive at very different conclusions. As you progress through the experiments in this book, you will enjoy genetics far more if you try to develop your own skills at formulating hypotheses, designing experiments, and interpreting data. Also, some of the questions in the problem sets are aimed at refining these skills.

Finally, it is worthwhile to point out that science is a social discipline. As you develop your skills at scrutinizing experiments, it is fun to discuss your ideas with other people, including fellow students and faculty members. Keep in mind that you do not need to "know all the answers" before you enter into a scientific discussion. Instead, it is more rewarding to view science as an ongoing and never-ending dialogue.

1.3 REVIEWING THE KEY CONCEPTS

- Model organisms are studied by many different researchers so they can compare their results and determine scientific principles that apply more broadly to other species (see Figure 1.13).
- Genetics is traditionally divided into transmission genetics, molecular genetics, and population genetics, though overlap occurs among these fields.
- Researchers in genetics carry out hypothesis testing or discovery-based science.

1.3 COMPREHENSION QUESTIONS

1. Which of the following is *not* a model organism?
   a. *Mus musculus* (laboratory mouse)
   b. *Escherichia coli* (a bacterium)
   c. *Saccharomyces cerevisiae* (a yeast)
   d. *Sciurus carolinensis* (gray squirrel)
2. A person studying the rate of transcription of a particular gene is working in the field of
   a. molecular genetics.
   b. transmission genetics.
   c. population genetics.
   d. None of the above.
3. The scientific method involves which of the following?
   a. The collection of observations and the formulation of a hypothesis
   b. Experimentation
   c. Data analysis and interpretation
   d. All of the above

KEY TERMS

Page 1. genome, deoxyribonucleic acid (DNA)
Page 4. genetics, gene, traits, nucleic acids, proteins, carbohydrates, lipids, macromolecules, proteome
Page 5. enzymes, nucleotides, polypeptides, genetic code, amino acid
Page 6. chromosomes, gene expression, transcription, ribonucleic acid (RNA), messenger RNA (mRNA), translation
Page 7. morphological traits, physiological traits, behavioral traits, molecular level, cellular level, organism level, species, population level, alleles
Page 8. genetic variation, morphs, gene mutations
Page 9. environment, norm of reaction, phenylketonuria (PKU)
Page 10. diploid, homologs, somatic cells, gametes, haploid, biological evolution, evolution, natural selection
Page 12. model organisms
Page 13. genetic cross
Page 14. genetic approach, loss-of-function mutation, loss-of-function allele
Page 15. hypothesis testing, scientific method, discovery-based science

CHAPTER SUMMARY

- The complete genetic composition of a cell or organism is called a genome. The genome encodes all of the proteins a cell or organism can make. Many key discoveries in genetics are related to the study of genes and genomes (see Figures 1.1–1.3).

1.1 The Molecular Expression of Genes

- Living cells are composed of nucleic acids (DNA and RNA), proteins, carbohydrates, and lipids. The proteome largely determines the structure and function of cells (see Figure 1.4).
- DNA, which is found within chromosomes, stores the information to make proteins (see Figure 1.5).
- Most genes encode polypeptides that are units within functional proteins. Gene expression at the molecular level involves transcription to produce mRNA and translation to produce a polypeptide (see Figure 1.6).

1.2 The Relationship Between Genes and Traits

- Genetics, which governs an organism's traits, spans the molecular, cellular, organism, and population levels (see Figure 1.7).
- Genetic variation underlies variation in traits. In addition, the environment plays a key role (see Figures 1.8–1.10).
- During reproduction, genetic material is passed from parents to offspring. In many species, somatic cells are diploid and have two sets of chromosomes whereas gametes are haploid and have a single set (see Figure 1.11).
- Evolution refers to a change in the genetic composition of a population from one generation to the next (see Figure 1.12).

1.3 Fields of Genetics

- Model organisms are studied by many different researchers so they can compare their results and determine scientific principles that apply more broadly to other species (see Figure 1.13).
- Genetics is traditionally divided into transmission genetics, molecular genetics, and population genetics, though overlap occurs among these fields.
- Researchers in genetics carry out hypothesis testing or discovery-based science.

PROBLEM SETS & INSIGHTS

Solved Problems

S1. A human gene called the *CFTR* gene (for cystic fibrosis transmembrane regulator) encodes a protein that functions in the transport of chloride ions across the cell membrane. Most people have two copies of a functional *CFTR* gene and do not have cystic fibrosis. However, a mutant version of the *CFTR* gene is found in some people. If a person has two mutant copies of the gene, he or she develops the disease known as cystic fibrosis. Are the following examples a description of genetics at the molecular, cellular, organism, or population level?

A. People with cystic fibrosis have lung problems due to a buildup of thick mucus in their lungs.

B. The mutant *CFTR* gene encodes a defective chloride transporter.

C. A defect in the chloride transporter causes a salt imbalance in lung cells.

D. Scientists have wondered why the mutant *CFTR* gene is relatively common. In fact, it is the most common mutant gene that causes a severe disease in Caucasians. Usually, mutant genes that cause severe diseases are relatively rare. One possible explanation why CF is so common is that people who have one copy of the functional *CFTR* gene and one copy of the mutant gene may be more resistant to diarrheal diseases such as cholera. Therefore, even though individuals with two mutant copies are very sick, people with one mutant copy and one functional copy might have a survival advantage over people with two functional copies of the gene.

Answer:

A. Organism. This is a description of a trait at the level of an entire individual.

B. Molecular. This is a description of a gene and the protein it encodes.

C. Cellular. This is a description of how protein function affects the cell.

D. Population. This is a possible explanation why two alleles of the gene occur within a population.

S2. Explain the relationship between the following pairs of terms:

A. RNA and DNA

B. RNA and transcription

C. Gene expression and trait

D. Mutation and allele

Answer:

A. DNA is the genetic material. DNA is used to make RNA. RNA is then used to specify a sequence of amino acids within a polypeptide.

B. Transcription is a process in which RNA is made using DNA as a template.

C. Genes are expressed at the molecular level to produce functional proteins. The functioning of proteins within living cells ultimately affects an organism's traits.

D. Alleles are alternative forms of the same gene. For example, a particular human gene affects eye color. The gene can exist as a blue allele or a brown allele. The difference between these two alleles is caused by a mutation. Perhaps the brown allele was the first eye color allele in the human population. Within some ancestral person, however, a mutation may have occurred in the eye color gene that converted the brown allele to the blue allele. Now the human population has both the brown allele and the blue allele.

S3. In diploid species that carry out sexual reproduction, how are genes passed from generation to generation?

Answer: When a diploid individual makes haploid cells for sexual reproduction, the cells contain half the number of chromosomes. When two haploid cells (e.g., sperm and egg) combine with each other, a zygote is formed that begins the life of a new individual. This zygote has inherited half of its chromosomes and, therefore, half of its genes from each parent. This is how genes are passed from parents to offspring.

Conceptual Questions

C1. At the molecular level, what is a gene? Where are genes located?

C2. Most genes encode proteins. Explain how the structure and function of proteins produce an organism's traits.

C3. Briefly explain how gene expression occurs at the molecular level.

C4. A human gene called the β-globin gene encodes a polypeptide that functions as a subunit of the protein known as hemoglobin. Hemoglobin carries oxygen within red blood cells. In human populations, the β-globin gene can be found as the more common allele called the Hb^A allele, but it can also be found as the Hb^S allele. Individuals who have two copies of the Hb^S allele have the disease called sickle cell disease. Are the following examples a description of genetics at the molecular, cellular, organism, or population level?

A. The Hb^S allele encodes a polypeptide that functions slightly differently from the polypeptide encoded by the Hb^A allele.

B. If an individual has two copies of the Hb^S allele, that person's red blood cells take on a sickle shape.

C. Individuals who have two copies of the Hb^A allele do not have sickle cell disease, but they are not resistant to malaria. People who have one Hb^A allele and one Hb^S allele do not have sickle cell disease, and they are resistant to malaria. People who have two copies of the Hb^S allele have sickle cell disease, and this disease may significantly shorten their lives.

D. Individuals with sickle cell disease have anemia because their red blood cells are easily destroyed by the body.

C5. What is meant by the term "genetic variation"? Give two examples of genetic variation not discussed in Chapter 1. What causes genetic variation at the molecular level?

C6. What is the cause of Down syndrome?

C7. Your textbook describes how the trait of phenylketonuria (PKU) is greatly influenced by the environment. Pick a trait in your favorite plant and explain how genetics and the environment may play important roles.

C8. What is meant by the term "diploid"? Which cells of the human body are diploid, and which cells are not?

C9. What is a DNA sequence?

C10. What is the genetic code?

C11. Explain the relationships between the following pairs of genetic terms:

A. Gene and trait

B. Gene and chromosome

C. Allele and gene

D. DNA sequence and amino acid sequence

C12. With regard to biological evolution, which of the following statements is incorrect? Explain why.

A. During its lifetime, an animal evolves to become better adapted to its environment.

B. The process of biological evolution has produced species that are better adapted to their environments.

C. When an animal is better adapted to its environment, the process of natural selection makes it more likely for that animal to reproduce.

C13. What are the primary interests of researchers working in the following fields of genetics?

A. Transmission genetics

B. Molecular genetics

C. Population genetics

Application and Experimental Questions

E1. Pick any example of a genetic technology and describe how it has directly affected your life.

E2. What is a genetic cross?

E3. The technique known as DNA sequencing (described in Chapter 19) enables researchers to determine the DNA sequence of genes. Would this technique be used primarily by transmission geneticists, molecular geneticists, or population geneticists?

E4. Figure 1.5 shows a micrograph of chromosomes from a normal human cell. If you performed this type of experiment using cells from a person with Down syndrome, what would you expect to see?

E5. Many organisms are studied by geneticists. Of the following species, do you think it is more likely for them to be studied by a transmission geneticist, a molecular geneticist, or a population geneticist? Explain your answer. Note: More than one answer may be possible.

A. Dogs
B. *E. coli*
C. Fruit flies
D. Leopards
E. Corn

E6. Pick any trait you like in any species of wild plant or animal. The trait must somehow vary among different members of the species. For example, some butterflies have dark wings and others have light wings (see Figure 1.7).

A. Discuss all of the background information that you already have (from personal observations) regarding this trait.

B. Propose a hypothesis that explains the genetic variation within the species. For example, in the case of the butterflies, your hypothesis might be that the dark butterflies survive better in dark forests, and the light butterflies survive better in sunlit fields.

C. Describe the experimental steps you would follow to test your hypothesis.

D. Describe the possible data you might collect.

E. Interpret your data.

Note: When picking a trait to answer this question, do not pick the trait of wing color in butterflies.

Answers to Comprehension Questions

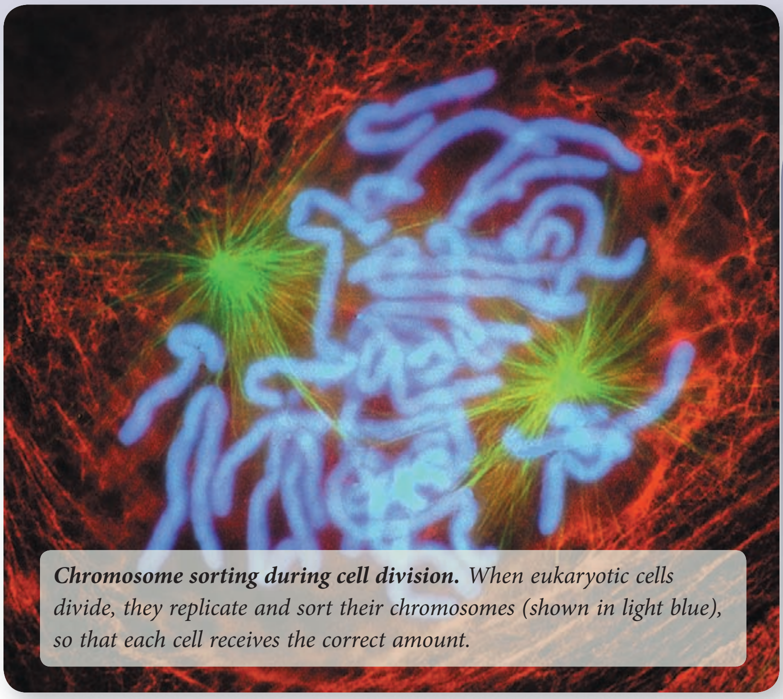
1.1: d, c, b, b
1.2: d, d, a, c
1.3: d, a, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

2



Chromosome sorting during cell division. *When eukaryotic cells divide, they replicate and sort their chromosomes (shown in light blue), so that each cell receives the correct amount.*

CHAPTER OUTLINE

2.1 General Features of Chromosomes
2.2 Cell Division
2.3 Mitosis and Cytokinesis
2.4 Meiosis
2.5 Sexual Reproduction

REPRODUCTION AND CHROMOSOME TRANSMISSION

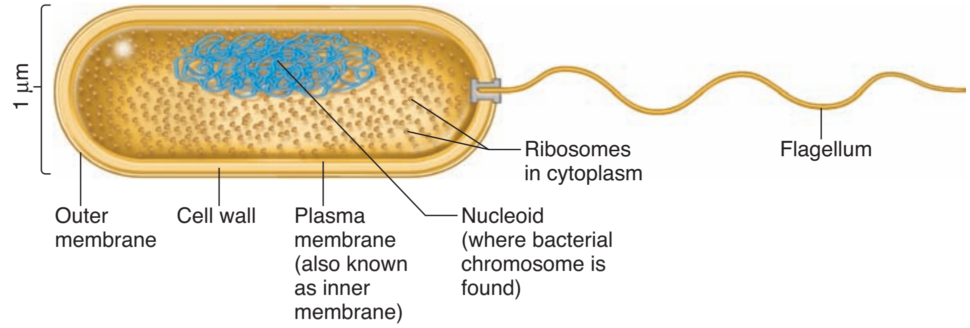
Reproduction is the biological process by which new cells or new organisms are produced. In this chapter, we will first survey reproduction at the cellular level, paying close attention to the inheritance of chromosomes. An examination of chromosomes at the microscopic level provides us with insights into understanding the inheritance patterns of traits, which we will consider in Chapter 3. To appreciate this relationship, we will examine how cells distribute their chromosomes during the process of cell division. We will see that in bacteria and most unicellular eukaryotes, simple cell division provides a way to reproduce asexually. Then we will explore a form of cell division called meiosis that produces cells with half the number of chromosomes. This form of cell division is needed for sexual reproduction, which is the formation of a new individual following the union of two gametes. This chapter will end with a discussion of how sexual reproduction occurs in animals and plants.

2.1 GENERAL FEATURES OF CHROMOSOMES

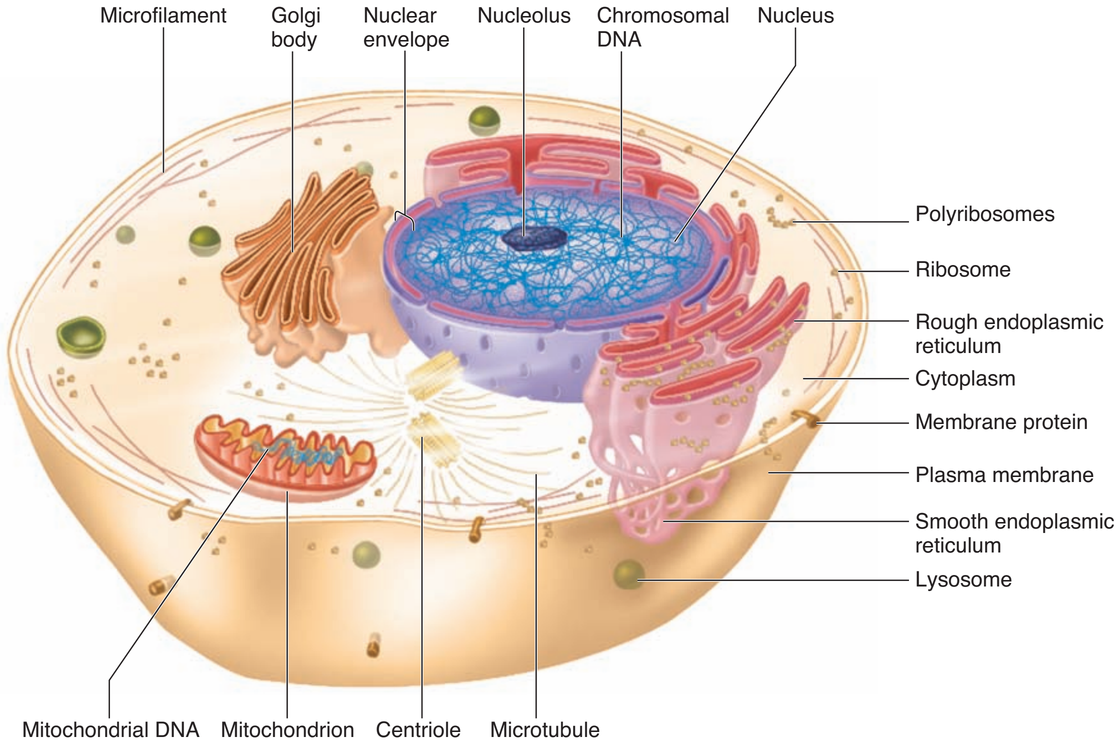
Chromosomes are structures within living cells that contain the genetic material. Genes are physically located within chromosomes. Biochemically, each chromosome contains a very long segment of DNA, which is the genetic material, and proteins, which are bound to the DNA and provide it with an organized structure. In eukaryotic cells, this complex between DNA and proteins is called **chromatin.** In this chapter, we will focus on the cellular mechanics of chromosome transmission to better understand the patterns of gene transmission that we will consider in Chapters 3 through 7. In particular, we will examine how chromosomes are copied and sorted into newly made cells. In later chapters, particularly Chapters 11 and 12, we will examine the molecular features of chromosomes in greater detail.

Before we begin a description of chromosome transmission, we need to consider the distinctive cellular differences between prokaryotic and eukaryotic species. Bacteria and Archaea are referred to as **prokaryotes,** from the Greek meaning prenucleus, because their chromosomes are not contained within a membrane-bound nucleus of the cell. Prokaryotes usually have a single type of circular chromosome in a region of the cytoplasm called the **nucleoid** (**Figure 2.1a**). The cytoplasm is enclosed by a plasma membrane that regulates the uptake of nutrients and the excretion of waste products. Outside the plasma membrane is a rigid cell wall that protects the cell from breakage. Certain species of bacteria also have an outer membrane located beyond the cell wall.

Eukaryotes, from the Greek meaning true nucleus, include some simple species, such as single-celled protists and some fungi (such as yeast), and more complex multicellular species, such as plants, animals, and other fungi. The cells of eukaryotic species have internal membranes that enclose highly specialized compartments (**Figure 2.1b**). These compartments form membrane-bound **organelles** with specific functions. For example, the lysosomes play a role in the degradation of macromolecules. The endoplasmic reticulum and Golgi body play a role in protein modification and trafficking. A particularly conspicuous organelle is the **nucleus,** which is bounded by two membranes that constitute the nuclear envelope. Most of the genetic material is found within chromosomes, which are located in the nucleus. In addition to the nucleus, certain organelles in eukaryotic cells contain a small amount of their own DNA. These include the mitochondrion, which plays a role in ATP synthesis, and, in plant cells, the chloroplast, which functions in photosynthesis.


1 µm
Ribosomes in cytoplasm
Flagellum
Outer membrane
Cell wall
Plasma membrane (also known as inner membrane)
Nucleoid (where bacterial chromosome is found)

(a) Bacterial cell


Microfilament
Golgi body
Nuclear envelope
Nucleolus
Chromosomal DNA
Nucleus
Polyribosomes
Ribosome
Rough endoplasmic reticulum
Cytoplasm
Membrane protein
Plasma membrane
Smooth endoplasmic reticulum
Lysosome
Mitochondrial DNA
Mitochondrion
Centriole
Microtubule

(b) Animal cell

FIGURE 2.1 **The basic organization of cells.** **(a)** A bacterial cell. The example shown here is typical of a bacterium such as *Escherichia coli*, which has an outer membrane. **(b)** A eukaryotic cell. The example shown here is a typical animal cell.

Concept check: *Eukaryotic cells exhibit compartmentalization. What does this mean?*

The DNA found in these organelles is referred to as extranuclear, or extrachromosomal, DNA to distinguish it from the DNA that is found in the cell nucleus. We will examine the role of mitochondrial and chloroplast DNA in Chapter 6.

In this section, we will focus on the composition of chromosomes found in the nucleus of eukaryotic cells. As you will learn, eukaryotic species contain genetic material that comes in sets of linear chromosomes.

Eukaryotic Chromosomes Are Examined Cytologically to Yield a Karyotype

Insights into inheritance patterns have been gained by observing chromosomes under the microscope. **Cytogenetics** is the field of genetics that involves the microscopic examination of chromosomes. The most basic observation that a **cytogeneticist** can make is to examine the chromosomal composition of a particular cell. For eukaryotic species, this is usually accomplished by observing the chromosomes as they are found in actively dividing cells. When a cell is preparing to divide, the chromosomes become more tightly coiled, which shortens them, thereby increasing their diameter. The consequence of this shortening is that distinctive shapes and numbers of chromosomes become visible with a light microscope. Each species has a particular chromosome composition. For example, most human cells contain 23 pairs of chromosomes, for a total of 46. On rare occasions, some individuals may inherit an abnormal number of chromosomes or a chromosome with an abnormal structure. Such abnormalities can often be detected by a microscopic examination of the chromosomes within actively dividing cells. In addition, a cytogeneticist may examine chromosomes as a way to distinguish between two closely related species.

Figure 2.2a shows the general procedure for preparing human chromosomes to be viewed by microscopy. In this example, the cells were obtained from a sample of human blood; more specifically, the chromosomes within leukocytes (a type of white blood cell) were examined. Blood cells are a type of **somatic cell,** which is any cell of the body that is not a gamete or a precursor to a gamete. Human **gametes** are sperm and egg cells. The gametes or their precursors are also called **germ cells.**

After the blood cells have been removed from the body, they are treated with drugs that stimulate them to begin cell division and cause cell division to be halted during mitosis, which is described later in this chapter. As shown in Figure 2.2a, these actively dividing cells are subjected to centrifugation to concentrate them. The concentrated preparation is then mixed with a hypotonic solution that makes the cells swell. This swelling causes the chromosomes to spread out within the cell, thereby making it easier to see each individual chromosome. Next, the cells are treated with a fixative that chemically freezes them so the chromosomes can no longer move around. The cells are then treated with a chemical dye that binds to the chromosomes and stains them. As discussed in greater detail in Chapter 8, this gives chromosomes a distinctive banding pattern that greatly enhances our ability to visualize and uniquely identify them (also see Figure 8.1c, d). The cells are then placed on a slide and viewed with a light microscope.

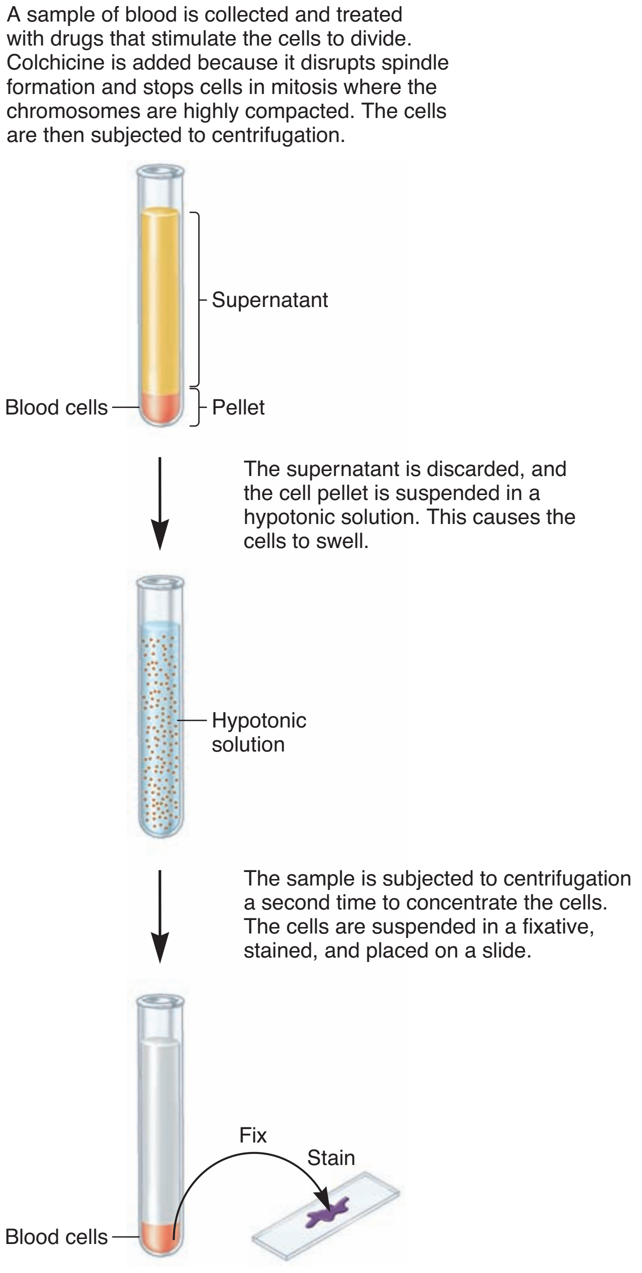
In a cytogenetics laboratory, the microscopes are equipped with a camera that can photograph the chromosomes. In recent years, advances in technology have allowed cytogeneticists to view microscopic images on a computer screen (**Figure 2.2b**). On the computer screen, the chromosomes can be organized in a standard way, usually from largest to smallest. As seen in **Figure 2.2c**, the human chromosomes have been lined up, and a number is given to designate each type of chromosome. An exception would be the sex chromosomes, which are designated with the letters X and Y. An organized representation of the chromosomes within a cell is called a **karyotype.** A karyotype reveals how many chromosomes are found within an actively dividing somatic cell.

Eukaryotic Chromosomes Are Inherited in Sets

Most eukaryotic species are **diploid** or have a diploid phase to their life cycles, which means that each type of chromosome is a member of a pair. A diploid cell has two sets of chromosomes. In humans, for example, most somatic cells have 46 chromosomes—two sets of 23 each. Other diploid species, however, can have different numbers of chromosomes in their somatic cells. For example, the dog has 39 chromosomes per set (78 total), the fruit fly has 4 chromosomes per set (8 total), and the tomato has 12 per set (24 total).

When a species is diploid, the members of a pair of chromosomes are called **homologs;** each type of chromosome is found in a homologous pair. As shown in Figure 2.2c, a human somatic cell has two copies of chromosome 1, two copies of chromosome 2, and so forth. Within each pair, the chromosome on the left is a homolog to the one on the right, and vice versa. In each pair, one chromosome was inherited from the mother and its homolog was inherited from the father. The two chromosomes in a homologous pair are nearly identical in size, have the same banding pattern, and contain a similar composition of genetic material. If a particular gene is found on one copy of a chromosome, it is also found on the other homolog. However, the two homologs may carry different versions of a given gene, which are called **alleles.** In Chapter 3, we will examine how some alleles are dominant, which mask the expression of alleles that are recessive. As an example, let's consider a gene in humans, called *OCA2,* which is one of a few different genes that affect eye color. The *OCA2* gene is located on chromosome 15 and comes in alleles that result in brown, green, or blue eyes. In a person with brown eyes, one copy of chromosome 15 might carry a dominant brown allele, whereas its homolog could carry a recessive blue allele. One copy was inherited from the mother and the other from the father.

At the molecular level, how similar are homologous chromosomes? The answer is that the sequence of bases of one homolog usually differs by less than 1% compared with the sequence of the other homolog. For example, the DNA sequence of chromosome 1 that you inherited from your mother is more than 99% identical to the sequence of chromosome 1 that you inherited from your father. Nevertheless, it should be emphasized that the sequences are not identical. The slight differences in DNA sequences provide the allelic differences in genes. Again, if we use the eye color gene as an example, a slight difference in



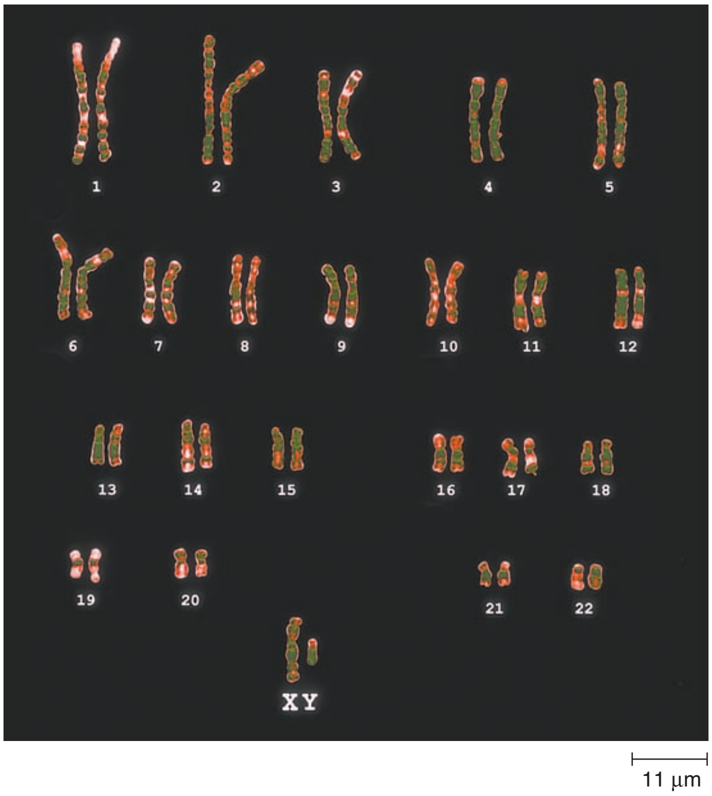
A sample of blood is collected and treated with drugs that stimulate the cells to divide. Colchicine is added because it disrupts spindle formation and stops cells in mitosis where the chromosomes are highly compacted. The cells are then subjected to centrifugation.

The supernatant is discarded, and the cell pellet is suspended in a hypotonic solution. This causes the cells to swell.

The sample is subjected to centrifugation a second time to concentrate the cells. The cells are suspended in a fixative, stained, and placed on a slide.

(a) Preparing cells for a karyotype

(b) The slide is viewed by a light microscope; the sample is seen on a video screen. The chromosomes can be arranged electronically on the screen.



(c) For a diploid human cell, two complete sets of chromosomes from a single cell constitute a karyotype of that cell.

FIGURE 2.2 **The procedure for making a human karyotype.**

Concept check: *How do you think the end results would be affected if the cells were not treated with a hypotonic solution?*

DNA sequence distinguishes the brown, green, and blue alleles. It should also be noted that the striking similarities between homologous chromosomes do not apply to the pair of sex chromosomes—X and Y. These chromosomes differ in size and genetic composition. Certain genes that are found on the X chromosome are not found on the Y chromosome, and vice versa. The X and Y chromosomes are not considered homologous chromosomes, though they do have short regions of homology.

Figure 2.3 considers two homologous chromosomes that are labeled with three different genes. An individual carrying

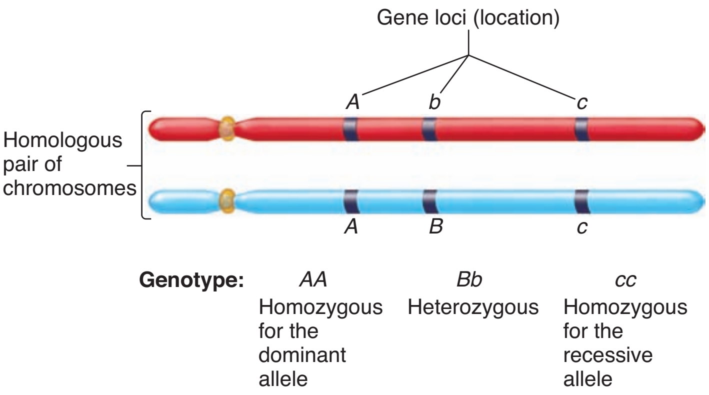

Gene loci (location)
A
b
c
Homologous pair of chromosomes
A
B
c
Genotype:
AA
Homozygous for the dominant allele
Bb
Heterozygous
cc
Homozygous for the recessive allele

FIGURE 2.3 **A comparison of homologous chromosomes.** Each pair of homologous chromosomes carries the same types of genes, but, as shown here, the alleles may or may not be different.

Concept check: How are homologs similar to each other and how are they different?

these two chromosomes is **homozygous** for the dominant allele of gene *A*, which means that both homologs carry the same allele. The individual is **heterozygous,** *Bb*, for the second gene, which means the homologs carry different alleles. For the third gene, the individual is homozygous for a recessive allele, *c*. The physical location of a gene is called its **locus** (plural: **loci**). As seen in Figure 2.3, for example, the locus of gene *C* is toward one end of this chromosome, whereas the locus of gene *B* is more in the middle.

2.1 REVIEWING THE KEY CONCEPTS

- Chromosomes are structures that contain the genetic material, which is DNA.
- Prokaryotic cells are simple and lack cell compartmentalization, but eukaryotic cells contain a cell nucleus and other compartments (see Figure 2.1).
- Chromosomes can be examined under the microscope. An organized representation of the chromosomes from a single cell is called a karyotype (see Figure 2.2).
- In eukaryotic species, the chromosomes are found in sets. Eukaryotic cells are often diploid, which means that each type of chromosome occurs in a homologous pair (see Figure 2.3).

2.1 COMPREHENSION QUESTIONS

1. Which of the following is not found in a prokaryotic cell?
   a. Plasma membrane
   b. Ribosome
   c. Cell nucleus
   d. Cytoplasm
2. When preparing a karyotype, which of the following steps is conducted?
   a. Treat the cells with drugs that cause them to begin cell division.
   b. Treat the cells with a hypotonic solution that causes them to swell.
   c. Expose the cells to chemical dyes that bind to the chromosomes and stain them.
   d. All of the above.
3. How many sets of chromosomes are found in a human somatic cell, and how many chromosomes are within one set?
   a. 2 sets, with 23 in each set.
   b. 23 sets, with 2 in each set.
   c. 1 set, with 23 in each set.
   d. 23 sets, with 1 in each set.

2.2 CELL DIVISION

Now that we have an appreciation for the chromosomal composition of living cells, we can consider how chromosomes are copied and transmitted when cells divide. One purpose of cell division is **asexual reproduction.** In this process, a preexisting cell divides to produce two new cells. By convention, the original cell is usually called the mother cell, and the new cells are the two daughter cells. When species are unicellular, the mother cell is judged to be one individual, and the two daughter cells are two new separate organisms. Asexual reproduction is how bacterial cells proliferate. In addition, certain unicellular eukaryotes, such as the amoeba and baker's yeast (*Saccharomyces cerevisiae*), can reproduce asexually.

A second important reason for cell division is multicellularity. Species such as plants, animals, most fungi, and some protists are derived from a single cell that has undergone repeated cellular divisions. Humans, for example, begin as a single fertilized egg; repeated cellular divisions produce an adult with many trillions of cells. The precise transmission of chromosomes during every cell division is critical so that all cells of the body receive the correct amount of genetic material.

In this section, we will consider how the process of cell division requires the duplication, organization, and distribution of the chromosomes. In bacteria, which have a single circular chromosome, the division process is relatively simple. Prior to cell division, bacteria duplicate their circular chromosome; they then distribute a copy into each of the two daughter cells. This process, known as binary fission, is described first. Eukaryotes have multiple numbers of chromosomes that occur as sets. Compared with bacteria, this added complexity requires a more complicated sorting process to ensure that each newly made cell receives the correct number and types of chromosomes. In this section, we will examine how eukaryotic cell division follows a series of stages called the cell cycle.

Bacteria Reproduce Asexually by Binary Fission

As discussed earlier (see Figure 2.1a), bacterial species are typically unicellular, although individual bacteria may associate with each other to form pairs, chains, or clumps. Unlike eukaryotes, which have their chromosomes in a separate nucleus, the circular chromosomes of bacteria are in direct contact with the cytoplasm. In Chapter 12, we will consider the molecular structure of bacterial chromosomes in greater detail.

The capacity of bacteria to divide is really quite astounding. Some species, such as *Escherichia coli,* a common bacterium of the intestine, can divide every 20 to 30 minutes. Prior to cell division, bacterial cells copy, or replicate, their chromosomal DNA. This produces two identical copies of the genetic material, as shown at the top of Figure 2.4. Following DNA replication, a bacterial cell divides into two daughter cells by a process known as **binary fission.** During this event, the two daughter cells become separated from each other by the formation of a partition called a septum. As seen in the figure, each cell receives a copy of the chromosomal genetic material. Except when rare mutations occur, the daughter cells are usually genetically identical because they contain exact copies of the genetic material from the mother cell.

Recent evidence has shown that bacterial species produce a protein called FtsZ, which is important in cell division. This protein assembles into a ring at the future site of the septum. FtsZ is thought to be the first protein to move to this division site, and it recruits other proteins that produce a new cell wall between the daughter cells. FtsZ is evolutionarily related to a eukaryotic protein called tubulin. As discussed later in this chapter, tubulin is the main component of microtubules, which play a key role in chromosome sorting in eukaryotes. Both FtsZ and tubulin form structures that provide cells with organization and play key roles in cell division.

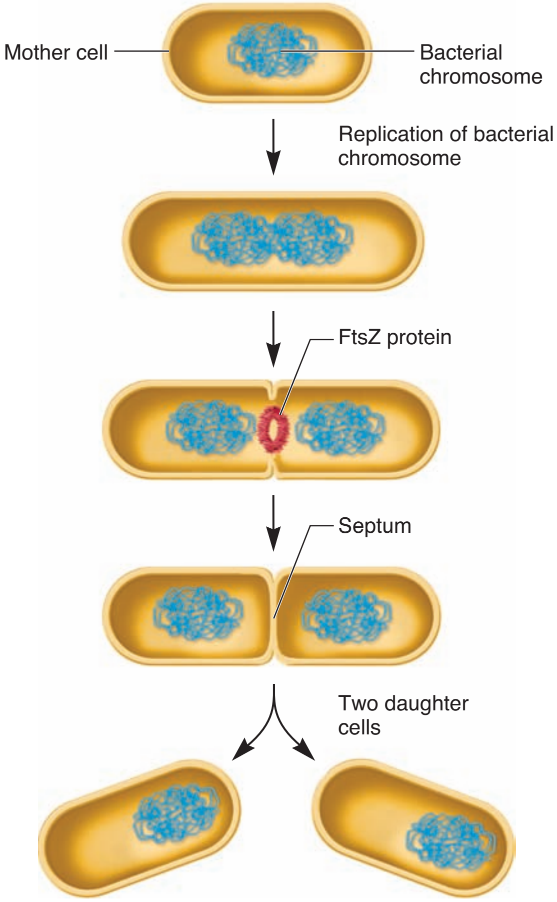


FIGURE 2.4 **Binary fission: The process by which bacterial cells divide.** Prior to division, the chromosome replicates to produce two identical copies. These two copies segregate from each other, with one copy going to each daughter cell.

Concept check: What is the function of the FtsZ protein during binary fission?

Binary fission is an asexual form of reproduction because it does not involve genetic contributions from two different gametes. On occasion, bacteria can exchange small pieces of genetic material with each other. We will consider some interesting mechanisms of genetic exchange in Chapter 9.

Eukaryotic Cells Progress Through a Cell Cycle to Produce Genetically Identical Daughter Cells

The common outcome of eukaryotic cell division is to produce two daughter cells that have the same number and types of chromosomes as the original mother cell. This requires a replication and division process that is more complicated than simple binary fission. Eukaryotic cells that are destined to divide progress through a series of phases known as the **cell cycle** (Figure 2.5). These phases are named G for gap, S for synthesis (of the genetic material), and M for mitosis. There are two G phases: G_1 and G_2. The term *gap* originally described the gaps between S phase and mitosis in which it was not microscopically apparent that significant changes were occurring in the cell. However, we now know that both gap phases are critical periods in the cell cycle that involve many molecular changes. In actively dividing cells, the G_1, S, and G_2 phases are collectively known as **interphase.** In addition, cells may remain permanently, or for long periods of time, in a phase of the cell cycle called G_0. A cell in the G_0 phase is either temporarily not progressing through the cell cycle or, in the case of terminally differentiated cells, such as most nerve cells in an adult mammal, will never divide again.

During the G_1 phase, a cell may prepare to divide. Depending on the cell type and the conditions it encounters, a cell in the G_1 phase may accumulate molecular changes (e.g., synthesis of proteins) that cause it to progress through the rest of the cell cycle. When this occurs, cell biologists say that a cell has reached a **restriction point** and is committed on a pathway that leads to cell division. Once past the restriction point, the cell then advances to the S phase, during which the chromosomes are replicated. After replication, the two copies are called **chromatids.** They are joined to each other at a region of DNA called the **centromere** to form a unit known as a pair of **sister chromatids** (Figure 2.6). The **kinetochore** is a group of proteins that are bound to the centromere. These proteins help to hold the sister chromatids together and also play a role in chromosome sorting, as discussed later. When S phase is completed, a cell actually has twice as many chromatids as chromosomes in the G_1 phase. For example, a human cell in the G_1 phase has 46 distinct chromosomes, whereas in G_2, it would have 46 pairs of sister chromatids, for a total of 92 chromatids. The term *chromosome,* which literally means "colored body," can be a bit confusing because it originally meant a distinct structure that is observable with the microscope. Therefore, the term *chromosome* can refer to either a pair of sister chromatids during the G_2 and early stages of M phase or to the structures that are observed at the end of M phase and those during G_1 that contain the equivalent of one chromatid (refer back to Figure 2.5).

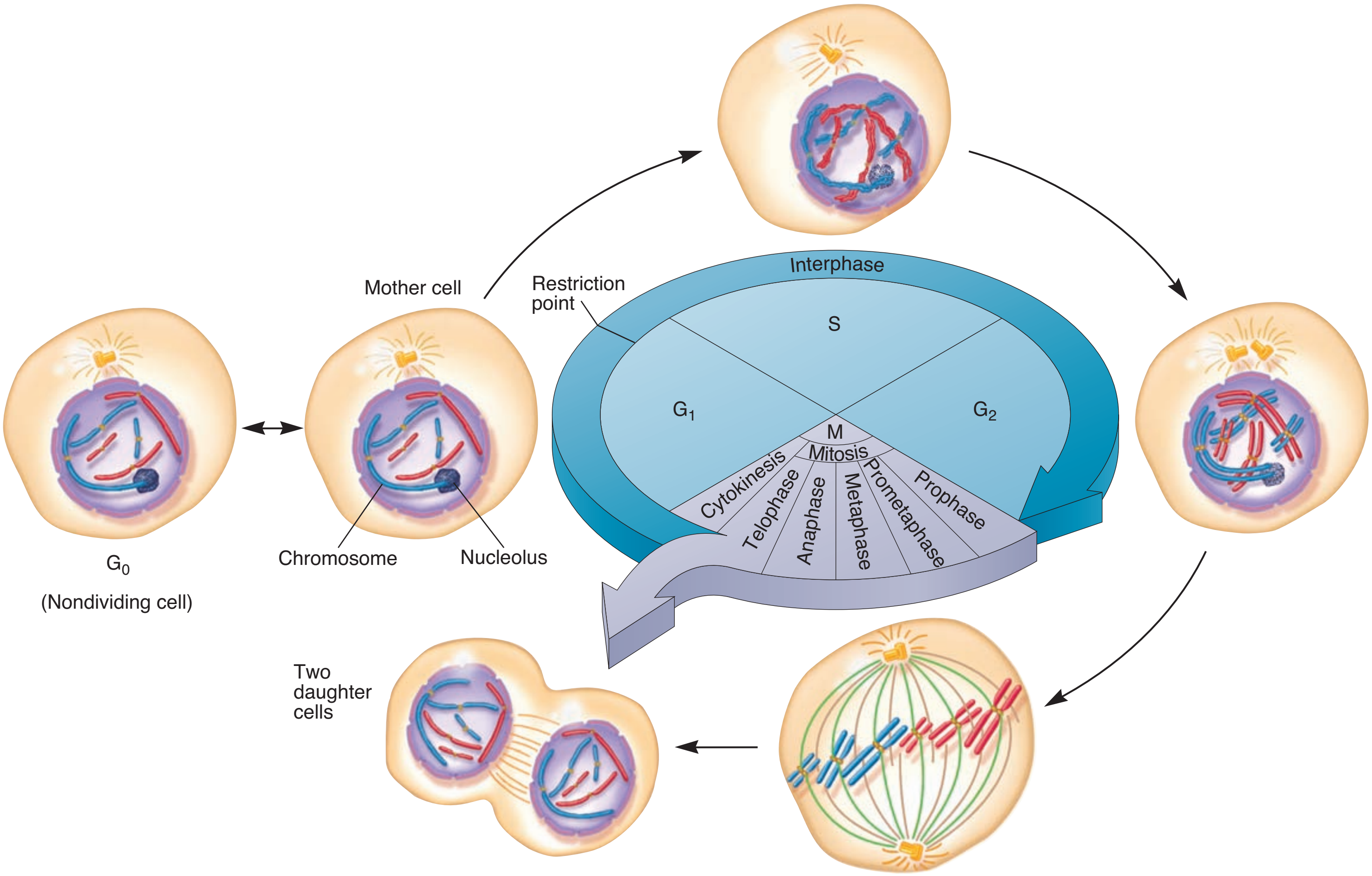


FIGURE 2.5 **The eukaryotic cell cycle.** Dividing cells progress through a series of phases, denoted G_1, S, G_2, and M phases. This diagram shows the progression of a cell through mitosis to produce two daughter cells. The original mother cell had three pairs of chromosomes, for a total of six individual chromosomes. By the G_2 phase, these have replicated to yield 12 chromatids found in six pairs of sister chromatids. After mitosis and cytokinesis are completed, each of the two daughter cells contains six individual chromosomes, just like the mother cell. Note: The chromosomes in G_0, G_1, S, and G_2 phases are not condensed (look ahead to Figure 2.8a). In this drawing, they are shown partially condensed so they can be easily counted.

Concept check: *What is the difference between the G_0 and G_1 phases?*

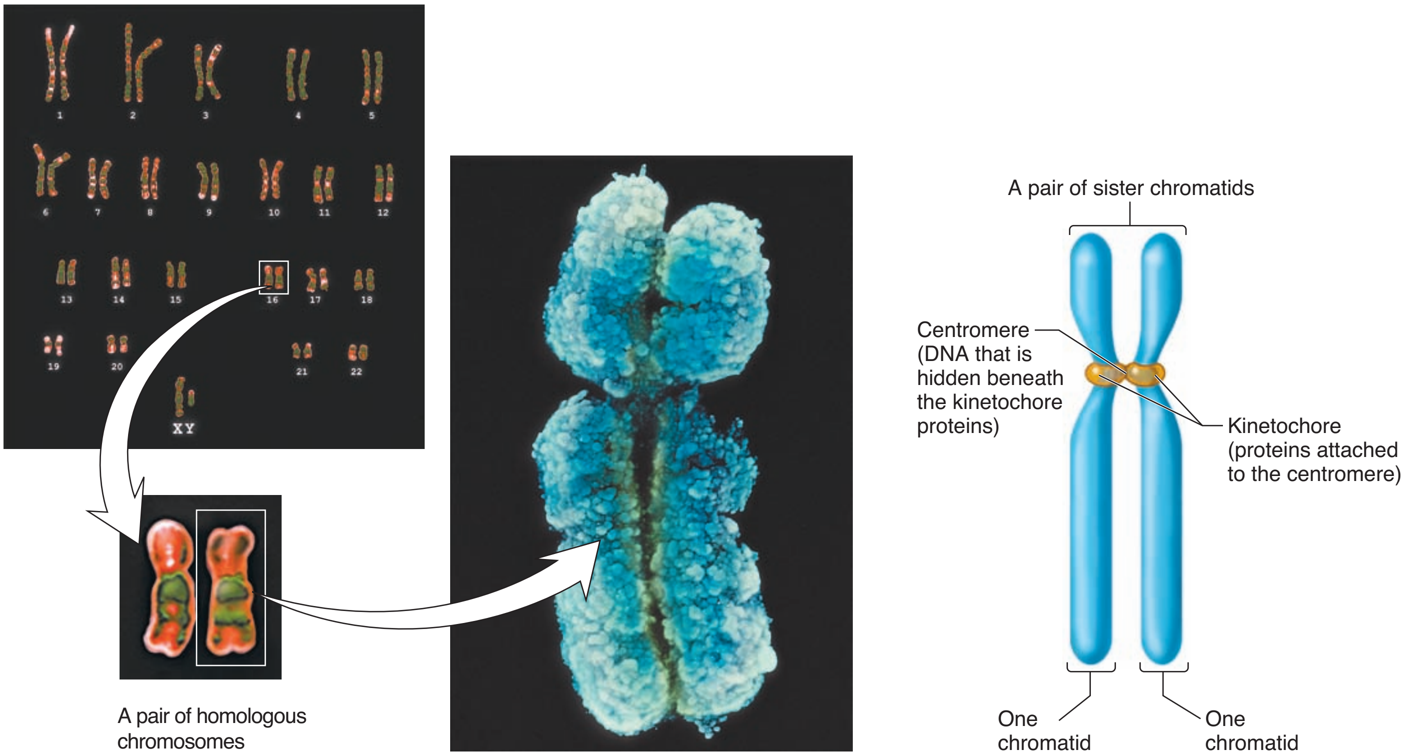
During the G_2 phase, the cell accumulates the materials that are necessary for nuclear and cell division. It then progresses into the M phase of the cell cycle, when **mitosis** occurs. The primary purpose of mitosis is to distribute the replicated chromosomes, dividing one cell nucleus into two nuclei, so that each daughter cell receives the same complement of chromosomes. For example, a human cell in the G_2 phase has 92 chromatids, which are found in 46 pairs. During mitosis, these pairs of chromatids are separated and sorted to produce two nuclei that contain 46 chromosomes each.

2.2 REVIEWING THE KEY CONCEPTS

- Bacteria divide by binary fission (see Figure 2.4).
- To divide, eukaryotic cells progress through a cell cycle (see Figure 2.5).
- Prior to cell division, eukaryotic chromosomes are replicated to form sister chromatids (see Figure 2.6).

2.2 COMPREHENSION QUESTIONS

1. Binary fission
   a. is a form of asexual reproduction.
   b. is a way for bacteria to reproduce.
   c. begins with a single mother cell and produces two genetically identical daughter cells.
   d. is all of the above.
2. Which of the following is the correct order of phases of the cell cycle?
   a. G_1, G_2, S, M
   b. G_1, S, G_2, M
   c. G_1, G_2, M, S
   d. G_1, S, M, G_2


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
XY
A pair of homologous chromosomes
A pair of sister chromatids
Centromere (DNA that is hidden beneath the kinetochore proteins)
Kinetochore (proteins attached to the centromere)
One chromatid
One chromatid

(a) Homologous chromosomes and sister chromatids

(b) Schematic drawing of sister chromatids

FIGURE 2.6 **Chromosomes following DNA replication.** **(a)** The photomicrograph on the right shows a chromosome in a form called a pair of sister chromatids. This chromosome is in the metaphase stage of mitosis, which is described later in the chapter. *Note:* Each of the 46 chromosomes that are viewed in a human karyotype (upper left) is actually a pair of sister chromatids. Look closely at the white rectangular boxes in the two insets. **(b)** A schematic drawing of sister chromatids. This structure has two chromatids that lie side by side. As seen here, each chromatid is a distinct unit. The two chromatids are held together by kinetochore proteins that bind to each other and to the centromeres of each chromatid.

Concept check: *What is the difference between homologs versus the chromatids within a pair of sister chromatids?*

3. What is a critical event that occurs during the S phase of the cell cycle?
   a. Cells make a decision of whether or not to divide.
   b. DNA replication produces pairs of sister chromatids.
   c. The chromosomes condense.
   d. The single nucleus is divided into two nuclei.

2.3 MITOSIS AND CYTOKINESIS

As we have seen, eukaryotic cell division involves a cell cycle in which the chromosomes are replicated and later sorted so that each daughter cell receives the same amount of genetic material. This process ensures genetic consistency from one cell generation to the next. In this section, we will examine the stages of mitosis and cytokinesis in greater detail.

The Mitotic Spindle Apparatus Organizes and Sorts Eukaryotic Chromosomes

Before we discuss the events of mitosis, let's first consider the structure of the **mitotic spindle apparatus** (also known simply as the **mitotic spindle**), which is involved in the organization and sorting of chromosomes (**Figure 2.7**). The spindle apparatus is formed from **microtubule-organizing centers** (MTOCs), which are structures found in eukaryotic cells from which microtubules grow. Microtubules are produced from the rapid polymerization of tubulin proteins. In animal cells, the mitotic spindle is formed from two MTOCs called **centrosomes.** Each centrosome is located at a **spindle pole.** A pair of **centrioles** at right angles to each other is found within each centrosome of animal cells. However, centrosomes and centrioles are not always found in eukaryotic species. For example, plant cells do not have centrosomes. Instead, the nuclear envelope functions as an MTOC for spindle formation.

The mitotic spindle of a typical animal cell has three types of microtubules (see Figure 2.7). The aster microtubules emanate outward from the centrosome toward the plasma membrane. They are important for the positioning of the spindle apparatus within the cell and later in the process of cell division. The polar microtubules project toward the region where the chromosomes are found during mitosis—the region between the two spindle poles. Polar microtubules that overlap with each other play a role in the separation of the two poles. They help to "push" the poles away from each other. Finally, the kinetochore microtubules have attachments

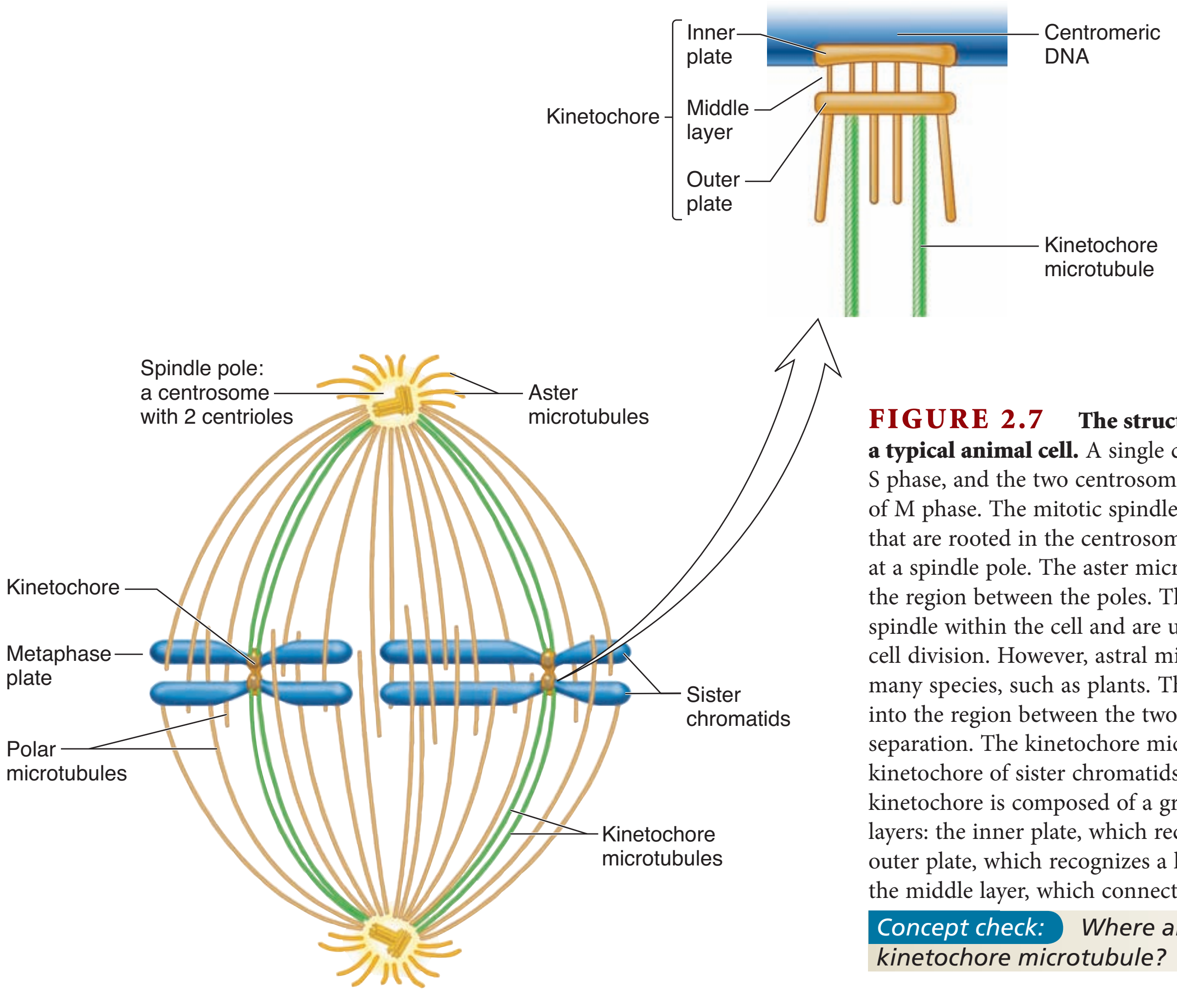


FIGURE 2.7 **The structure of the mitotic spindle in a typical animal cell.** A single centrosome duplicates during S phase, and the two centrosomes separate at the beginning of M phase. The mitotic spindle is formed from microtubules that are rooted in the centrosomes. Each centrosome is located at a spindle pole. The aster microtubules emanate away from the region between the poles. They help to position the spindle within the cell and are used as reference points for cell division. However, astral microtubules are not found in many species, such as plants. The polar microtubules project into the region between the two poles; they play a role in pole separation. The kinetochore microtubules are attached to the kinetochore of sister chromatids. As seen in the inset, the kinetochore is composed of a group of proteins that form three layers: the inner plate, which recognizes the centromere; the outer plate, which recognizes a kinetochore microtubule; and the middle layer, which connects the inner and outer plates.

Concept check: Where are the two ends of a kinetochore microtubule?

to a kinetochore, which is a complex of proteins that is bound to the centromere of individual chromosomes. As seen in the inset to Figure 2.7, the kinetochore proteins form three layers. The proteins of the inner plate make direct contact with the centromeric DNA, whereas the outer plate contacts the kinetochore microtubules. The role of the middle layer is to connect these two regions.

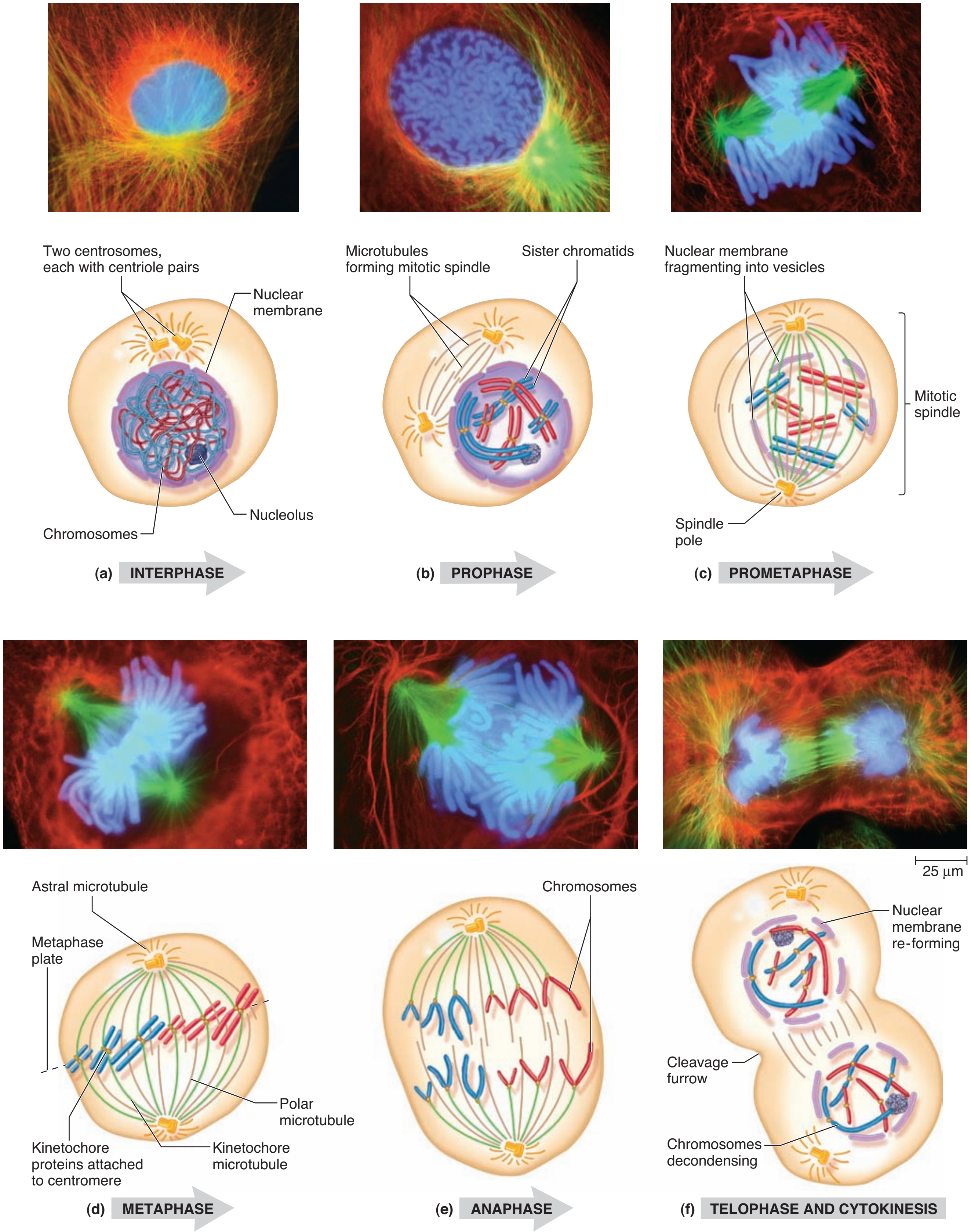
The mitotic spindle allows cells to organize and separate chromosomes so that each daughter cell receives the same complement of chromosomes. This sorting process, known as mitosis, is described next.

The Transmission of Chromosomes During the Division of Eukaryotic Cells Involves a Process Known as Mitosis

In Figure 2.8, the process of mitosis is shown for a diploid animal cell. In the simplified diagrams shown along the bottom of this figure, the original mother cell contains six chromosomes; it is diploid ($2n$) and contains three chromosomes per set ($n = 3$). One set is shown in blue, and the homologous set is shown in red. As discussed next, mitosis is subdivided into phases known as prophase, prometaphase, metaphase, anaphase, and telophase.

Prophase Prior to mitosis, the cells are in interphase, during which the chromosomes are **decondensed**—less tightly compacted—and found in the nucleus (Figure 2.8a). At the start of mitosis, in **prophase,** the chromosomes have already replicated to produce 12 chromatids, joined as six pairs of sister chromatids (Figure 2.8b). As prophase proceeds, the nuclear membrane begins to dissociate into small vesicles. At the same time, the chromatids **condense** into more compact structures that are readily visible by light microscopy. The mitotic spindle also begins to form and the nucleolus, which is the site of ribosome assembly, disappears.

Prometaphase As mitosis progresses from prophase to prometaphase, the centrosomes move to opposite ends of the cell and demarcate two spindle poles, one within each of the future daughter cells. Once the nuclear membrane has dissociated into vesicles, the spindle fibers can interact with the sister chromatids. This interaction occurs in a phase of mitosis called **prometaphase** (Figure 2.8c). How do sister chromatids become attached to the spindle? Initially, microtubules are rapidly formed and can be seen growing out from the two poles. As it grows, if the end of a microtubule happens to make contact with a kinetochore, its end


Two centrosomes, each with centriole pairs
Nuclear membrane
Chromosomes
Nucleolus
(a) INTERPHASE
Microtubules forming mitotic spindle
Sister chromatids
(b) PROPHASE
Nuclear membrane fragmenting into vesicles
Mitotic spindle
Spindle pole
(c) PROMETAPHASE
25 µm
Astral microtubule
Metaphase plate
Polar microtubule
Kinetochore proteins attached to centromere
Kinetochore microtubule
(d) METAPHASE
Chromosomes
(e) ANAPHASE
Nuclear membrane re-forming
Cleavage furrow
Chromosomes decondensing
(f) TELOPHASE AND CYTOKINESIS

ONLINE ANIMATION

FIGURE 2.8 **The process of mitosis in an animal cell.** The top panels illustrate cells of a fish embryo progressing through mitosis. The chromosomes are stained in blue and the spindle is green. The bottom panels are schematic drawings that emphasize the sorting and separation of the chromosomes. In this case, the original diploid cell had six chromosomes (three in each set). At the start of mitosis, these have already replicated into 12 chromatids. The final result is two daughter cells, each containing six chromosomes.

Concept check: *During which phase are sister chromatids separated and sent to opposite poles?*

is said to be "captured" and remains firmly attached to the kinetochore. This random process is how sister chromatids become attached to kinetochore microtubules. Alternatively, if the end of a microtubule does not collide with a kinetochore, the microtubule eventually depolymerizes and retracts to the centrosome. As the end of prometaphase nears, the kinetochore on a pair of sister chromatids is attached to kinetochore microtubules from opposite poles. As these events are occurring, the sister chromatids are seen to undergo jerky movements as they are tugged, back and forth, between the two poles. By the end of prometaphase, the mitotic spindle is completely formed.

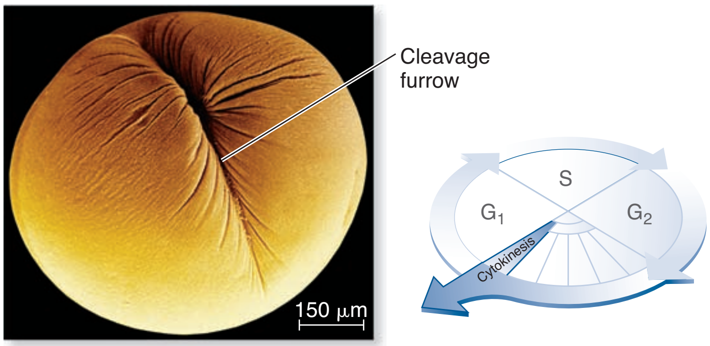
Metaphase Eventually, the pairs of sister chromatids align themselves along a plane called the **metaphase plate.** As shown in Figure 2.8d, when this alignment is complete, the cell is in **metaphase** of mitosis. At this point, each pair of chromatids is attached to both poles by kinetochore microtubules. The pairs of sister chromatids have become organized into a single row along the metaphase plate. When this organizational process is finished, the chromatids can be equally distributed into two daughter cells.

Anaphase At **anaphase,** the connection that is responsible for holding the pairs of chromatids together is broken (Figure 2.8e). (We will examine the process of sister chromatid cohesion and separation in more detail in Chapter 12.) Each chromatid, now an individual chromosome, is linked to only one of the two poles. As anaphase proceeds, the chromosomes move toward the pole to which they are attached. This involves a shortening of the kinetochore microtubules. In addition, the two poles themselves move farther apart due to the elongation of the polar microtubules, which slide in opposite directions due to the actions of motor proteins.

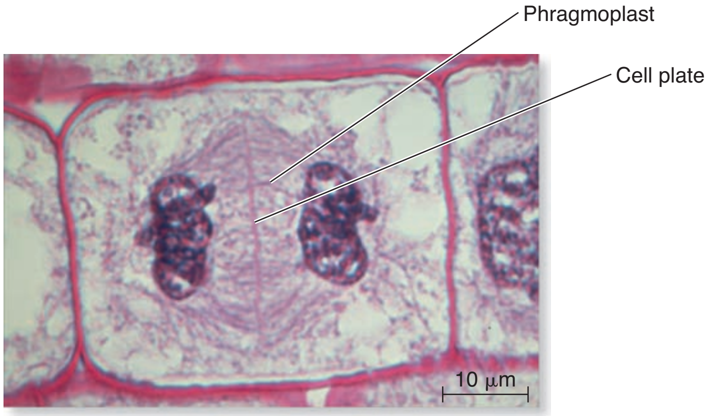
Telophase During **telophase,** the chromosomes reach their respective poles and decondense. The nuclear membrane now reforms to produce two separate nuclei. In Figure 2.8f, this has produced two nuclei that contain six chromosomes each. The nucleoli will also reappear.

Cytokinesis In most cases, mitosis is quickly followed by **cytokinesis,** in which the two nuclei are segregated into separate daughter cells. Likewise, cytokinesis also segregates cell organelles such as mitochondria and chloroplasts into daughter cells. In animal cells, cytokinesis begins shortly after anaphase. A contractile ring, composed of myosin motor proteins and actin filaments, assembles adjacent to the plasma membrane. Myosin hydrolyzes ATP, which shortens the ring and thereby constricts the plasma membrane to form a **cleavage furrow** that ingresses, or moves inward (**Figure 2.9a**). Ingression continues until a midbody structure is formed that physically pinches one cell into two.

In plants, the two daughter cells are separated by the formation of a **cell plate** (**Figure 2.9b**). At the end of anaphase, Golgi-derived vesicles carrying cell wall materials are transported to the equator of a dividing cell. These vesicles are directed to their location via the phragmoplast, which is composed of parallel aligned microtubules and actin filaments that serve as tracks for vesicle movement. The fusion of these vesicles gives rise to the cell plate, which is a membrane-bound compartment. The cell plate begins in the middle of the cell and expands until it attaches to the mother cell wall. Once this attachment has taken place, the cell plate undergoes a process of maturation and eventually separates the mother cell into two daughter cells.



(a) Cleavage of an animal cell



(b) Formation of a cell plate in a plant cell

FIGURE 2.9 **Cytokinesis in an animal and plant cell.** (**a**) In an animal cell, cytokinesis involves the formation of a cleavage furrow. (**b**) In a plant cell, cytokinesis occurs via the formation of a cell plate between the two daughter cells.

Concept check: *What causes the cleavage furrow to ingress?*

Outcome of Mitotic Cell Division Mitosis and cytokinesis ultimately produce two daughter cells having the same number of chromosomes as the mother cell. Barring rare mutations, the two daughter cells are genetically identical to each other and to the mother cell from which they were derived. The critical consequence of this sorting process is to ensure genetic consistency from one somatic cell to the next. The development of multicellularity relies on the repeated process of mitosis and cytokinesis. For diploid organisms that are multicellular, most of the somatic cells are diploid and genetically identical to each other.

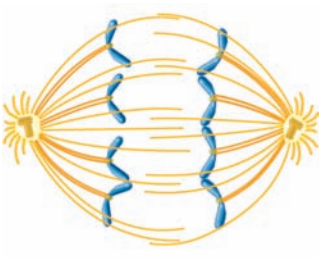
2.3 REVIEWING THE KEY CONCEPTS

- Chromosome sorting in eukaryotes is achieved via a spindle apparatus (see Figure 2.7).

- A common way for eukaryotic cells to divide is by mitosis and cytokinesis. Mitosis is divided into prophase, prometaphase, metaphase, anaphase, and telophase (see Figures 2.8, 2.9).

2.3 COMPREHENSION QUESTIONS

1. What is the function of microtubules during mitosis?
   a. Aster microtubules are important for the positioning of the spindle apparatus within an animal cell and in the process of cell division.
   b. Polar microtubules play a role in the separation of the two poles.
   c. Kinetochore microtubules aid in the alignment of chromosomes during metaphase and the movement of individual chromosomes toward a pole during anaphase.
   d. All of the above
2. Which phase of mitosis is depicted in the drawing below?



   a. Prophase
   b. Prometaphase
   c. Metaphase
   d. Anaphase
   e. Telophase

2.4 MEIOSIS

In the previous section, we considered how eukaryotic cells can divide by mitosis and cytokinesis so that a mother cell produces two genetically identical daughter cells. Diploid eukaryotic cells may also divide by an alternative process called **meiosis** (from the Greek meaning "less"). During meiosis, **haploid** cells, which contain a single set of chromosomes, are produced from a cell that was originally diploid. For this to occur, the chromosomes must be correctly sorted and distributed in a way that reduces the chromosome number to half its original value. In the case of humans, for example, each gamete must receive half of the total number of chromosomes. But not just any 23 chromosomes will do—a gamete must receive one chromosome from each of the 23 pairs. In this section, we examine how the stages of meiosis lead to the formation of cells with a haploid complement of chromosomes.

Meiosis Produces Cells That Are Haploid

The process of meiosis bears striking similarities to mitosis. Like mitosis, meiosis begins after a cell has progressed through the G_1, S, and G_2 phases of the cell cycle. However, meiosis involves two successive divisions rather than one (as in mitosis). Prior to meiosis, the chromosomes are replicated in S phase to produce pairs of sister chromatids. This single replication event is then followed by two sequential cell divisions: called meiosis I and II. As in mitosis, each of these is subdivided into prophase, prometaphase, metaphase, anaphase, and telophase.

Prophase of Meiosis I Figure 2.10 emphasizes some of the important events that occur during prophase of meiosis I, which is further subdivided into stages known as leptotene, zygotene, pachytene, diplotene, and diakinesis. During the **leptotene** stage, the replicated chromosomes begin to condense and become visible with a light microscope. Unlike mitosis, the **zygotene** stage of prophase of meiosis I involves a recognition process known as **synapsis,** in which the homologous chromosomes recognize each other and begin to align themselves. As this occurs, a **synaptonemal complex** forms between the homologs. At **pachytene,** the homologs have become completely aligned. The associated chromatids are known as **bivalents.** Each bivalent contains two pairs of sister chromatids, or a total of four chromatids. A bivalent is also called a **tetrad** (from the prefix *tetra-*, meaning "four") because it is composed of four chromatids.

Prior to the pachytene stage, when synapsis is complete, an event known as crossing over usually occurs. **Crossing over** involves a physical exchange of chromosome pieces that results in an exchange of genetic information. Depending on the size of the chromosome and the species, an average eukaryotic chromosome incurs a couple to a couple dozen crossovers. During spermatogenesis in humans, for example, an average chromosome undergoes slightly more than 2 crossovers, whereas chromosomes in certain plant species may undergo 20 or more crossovers. Recent research has shown that crossing over is usually critical for the proper segregation of chromosomes. In fact, abnormalities in chromosome segregation may be related to a defect in crossing over. In a high percentage of people with Down syndrome, in which an individual has three copies of chromosome 21 instead of two, research has shown that the presence of the extra chromosome is associated with a lack of crossing over between homologous chromosomes.

In Figure 2.10, crossing over has occurred at a single site between two of the larger chromatids. The connection that results from crossing over is called a **chiasma** (plural: **chiasmata**), because it physically resembles the Greek letter chi, χ. We will consider the genetic consequences of crossing over in Chapter 7 and the molecular process of crossing over in Chapter 13. By the end of the **diplotene** stage, the synaptonemal complex has largely disappeared. The bivalent pulls apart slightly, and microscopically it becomes easier to see that it is actually composed of four chromatids. In the last stage of prophase of meiosis I, **diakinesis,** the synaptonemal complex completely disappears.

As mentioned, a synaptonemal complex is formed between the homologous chromosomes, but this does not occur in all eukaryotic species. As shown in Figure 2.11, this complex is composed of parallel lateral elements, which are bound to the chromosomal DNA, and a central element, which promotes the binding of the lateral elements to each other via transverse filaments. The synaptonemal complex may not be required for the pairing of homologous chromosomes, because some species, such as *Aspergillus nidulans* and *Schizosaccharomyces pombe,* completely lack such a complex, yet their chromosomes synapse correctly. At present, the precise role of the synaptonemal

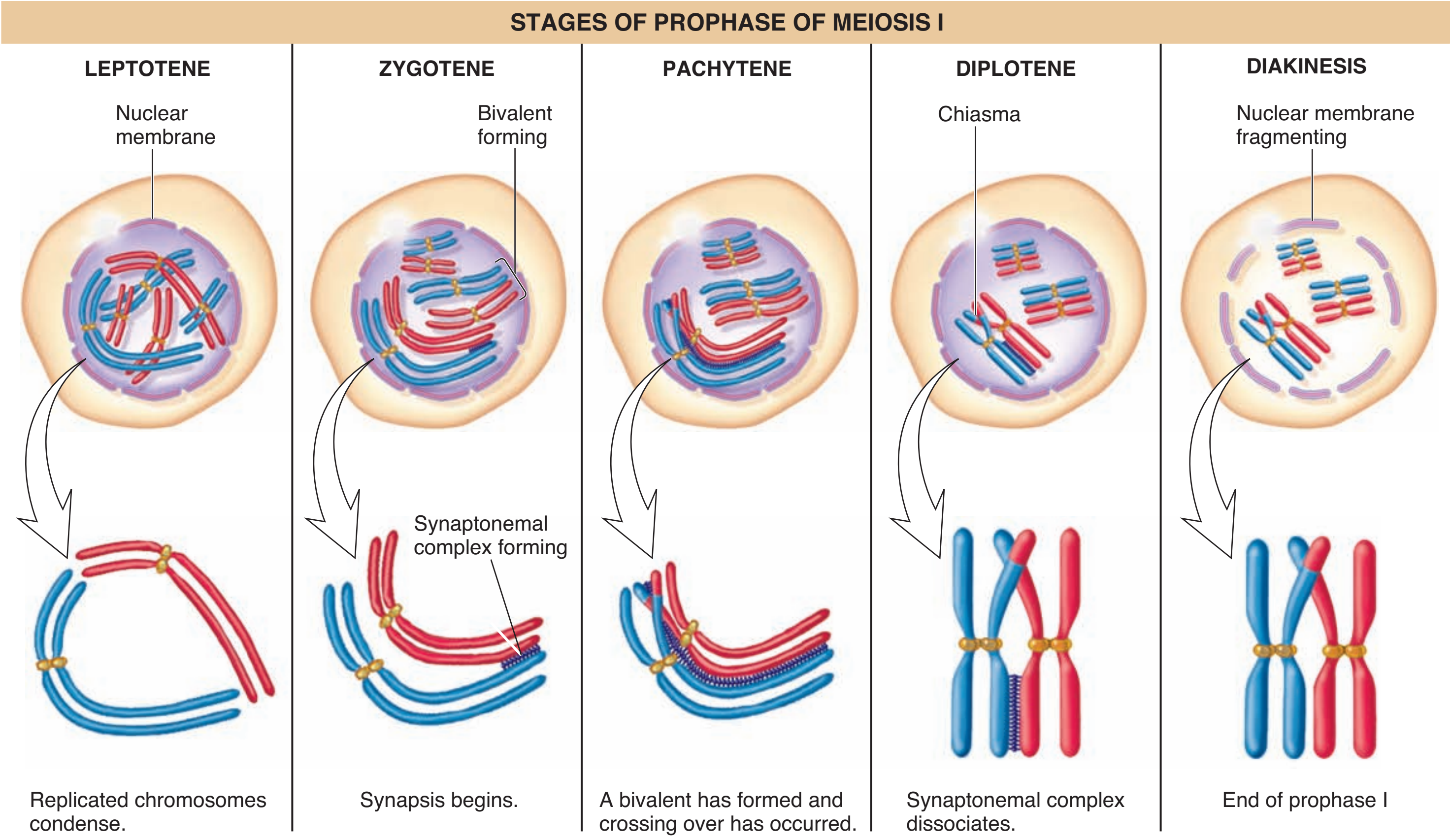

STAGES OF PROPHASE OF MEIOSIS I
LEPTOTENE
Nuclear membrane
Replicated chromosomes condense.
ZYGOTENE
Bivalent forming
Synaptonemal complex forming
Synapsis begins.
PACHYTENE
A bivalent has formed and crossing over has occurred.
DIPLOTENE
Chiasma
Synaptonemal complex dissociates.
DIAKINESIS
Nuclear membrane fragmenting
End of prophase I

FIGURE 2.10 **The events that occur during prophase of meiosis I.** *Note:* for simplicity, the nucleolus is not shown in this figure.

Concept check: *What is the end result of crossing over?*

complex is not clearly understood, and it remains the subject of research. It may play more than one role. First, although it may not be required for synapsis, the synaptonemal complex could help to maintain homologous pairing in situations where the normal process has failed. Second, the complex may play a role in meiotic chromosome structure. And third, the synaptonemal complex may serve to regulate the process of crossing over.

Prometaphase of Meiosis I Figure 2.10 emphasized the pairing and crossing over that occurs during prophase of meiosis I. In **Figure 2.12**, we turn our attention to the general events in meiosis. Prophase of meiosis I is followed by prometaphase, in which the spindle apparatus is complete and the chromatids are attached via kinetochore microtubules.

Metaphase of Meiosis I At metaphase of meiosis I, the bivalents are organized along the metaphase plate. However, their pattern of alignment is strikingly different from that observed during mitosis

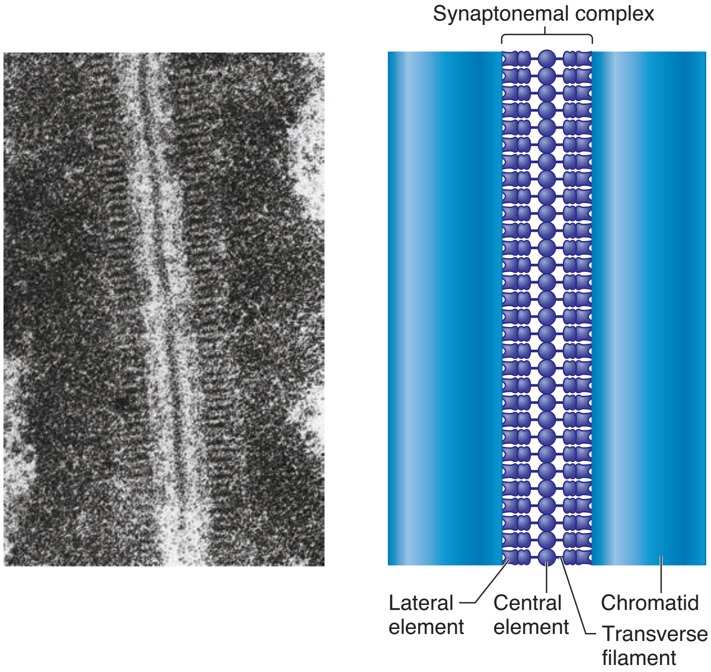
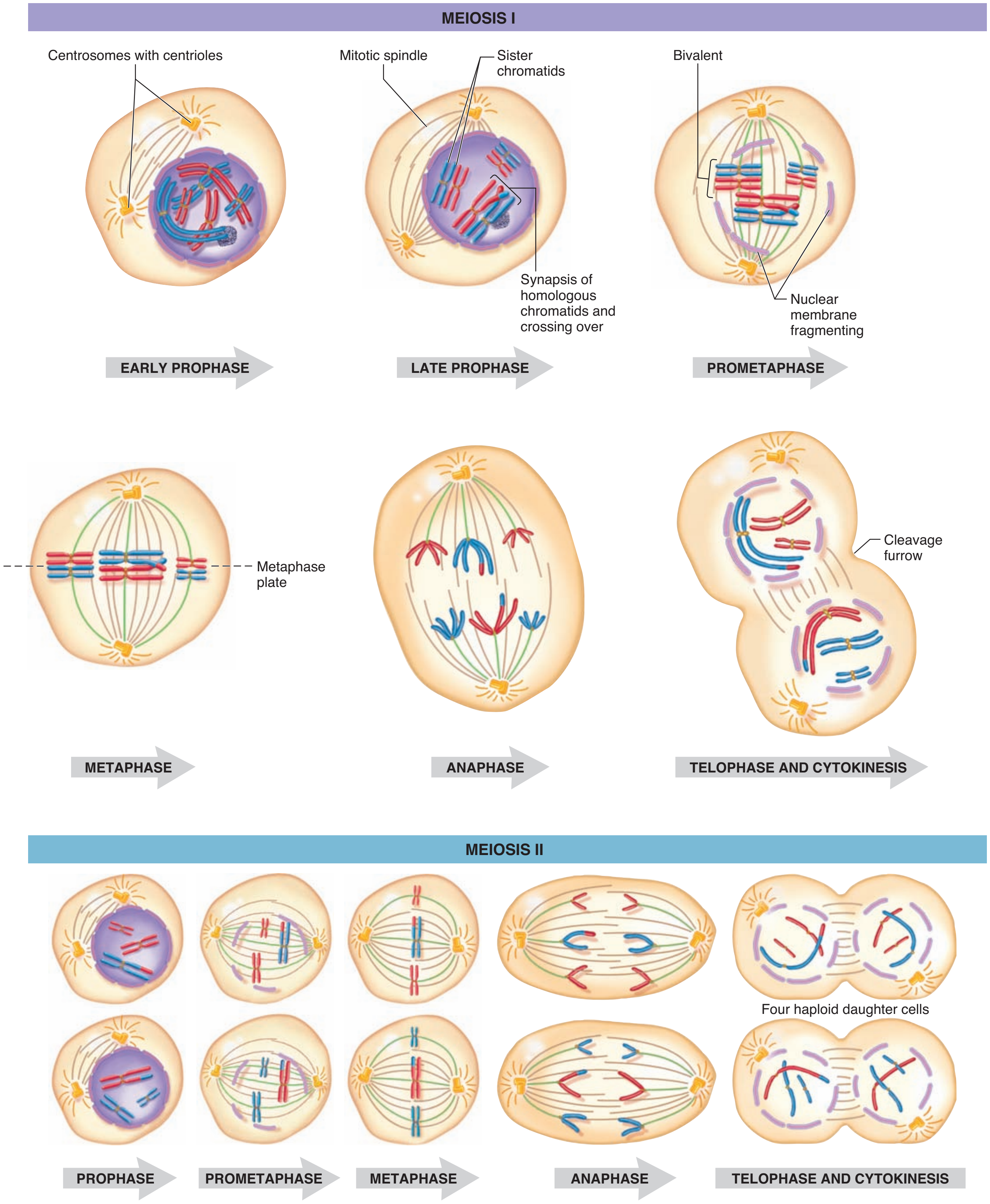

Synaptonemal complex
Lateral element
Central element
Chromatid
Transverse filament

FIGURE 2.11 **The synaptonemal complex formed during prophase of meiosis I.** The left side is a transmission electron micrograph of a synaptonemal complex. Lateral elements are bound to the chromosomal DNA of homologous chromatids. A central element provides a link between the lateral elements via transverse filaments.

Concept check: *Where is the synaptonemal complex located? Is it between homologs, or is it between the chromatids within a pair of sister chromatids?*


MEIOSIS I
Centrosomes with centrioles
Mitotic spindle
Sister chromatids
Bivalent
Synapsis of homologous chromatids and crossing over
Nuclear membrane fragmenting
EARLY PROPHASE
LATE PROPHASE
PROMETAPHASE
Metaphase plate
Cleavage furrow
METAPHASE
ANAPHASE
TELOPHASE AND CYTOKINESIS
MEIOSIS II
Four haploid daughter cells
PROPHASE
PROMETAPHASE
METAPHASE
ANAPHASE
TELOPHASE AND CYTOKINESIS

ONLINE ANIMATION

FIGURE 2.12 **The stages of meiosis in an animal cell.** See text for details.

Concept check: *How do the four cells at the end of meiosis differ from the original mother cell?*

(refer back to Figure 2.8d). Before we consider the rest of meiosis I, a particularly critical feature for you to appreciate is how the bivalents are aligned along the metaphase plate. In particular, the pairs of sister chromatids are aligned in a double row rather than a single row, as occurs in mitosis. Furthermore, the arrangement of sister chromatids within this double row is random with regard to the blue and red homologs. In Figure 2.12, one of the blue homologs is above the metaphase plate and the other two are below, whereas one of the red homologs is below the metaphase plate and other two are above.

In an organism that produces many gametes, meiosis in other cells can produce a different arrangement of homologs—three blues above and none below, or none above and three below, and so on. Because most eukaryotic species have several chromosomes per set, the sister chromatids can be randomly aligned along the metaphase plate in many possible ways. For example, consider humans, who have 23 chromosomes per set. The possible number of different, random alignments equals 2^n, where n equals the number of chromosomes per set. Thus, in humans, this equals 2^{23}, or over 8 million, possibilities! Because the homologs are genetically similar but not identical, we see from this calculation that the random alignment of homologous chromosomes provides a mechanism to promote a vast amount of genetic diversity.

In addition to the random arrangement of homologs within a double row, a second distinctive feature of metaphase of meiosis I is the attachment of kinetochore microtubules to the sister chromatids (**Figure 2.13**). One pair of sister chromatids is linked to one of the poles, and the homologous pair is linked to the opposite pole. This arrangement is quite different from the kinetochore attachment sites during mitosis in which a pair of sister chromatids is linked to both poles (see Figure 2.8d).

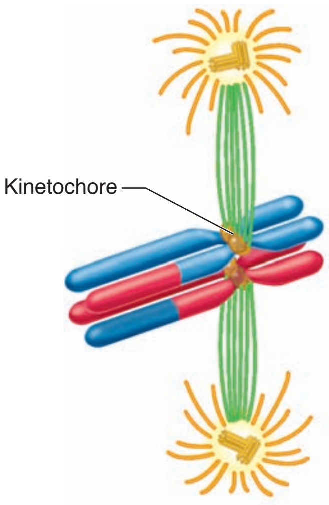


FIGURE 2.13 **Attachment of the kinetochore microtubules to replicated chromosomes during meiosis.** The kinetochore microtubules from a given pole are attached to one pair of chromatids in a bivalent, but not both. Therefore, each pair of sister chromatids is attached to only one pole.

Concept check: How is this attachment of chromosomes to kinetochore microtubules different from their attachment during metaphase of mitosis?

Anaphase of Meiosis I During anaphase of meiosis I, the two pairs of sister chromatids within a bivalent separate from each other (see Figure 2.12). However, the connection that holds sister chromatids together does not break. Instead, each joined pair of chromatids migrates to one pole, and the homologous pair of chromatids moves to the opposite pole.

Telophase of Meiosis I and Cytokinesis Finally, at telophase of meiosis I, the sister chromatids have reached their respective poles, and decondensation occurs in many, but not all, species. The nuclear membrane may re-form to produce two separate nuclei. The end result of meiosis I is two cells, each with three pairs of sister chromatids. It is a reduction division. The original diploid cell had its chromosomes in homologous pairs, but the two cells produced at the end of meiosis I are considered to be haploid; they do not have pairs of homologous chromosomes.

Meiosis II The sorting events that occur during meiosis II are similar to those that occur during mitosis, but the starting point is different. For a diploid organism with six chromosomes, mitosis begins with 12 chromatids that are joined as six pairs of sister chromatids (refer back to Figure 2.8). By comparison, the two cells that begin meiosis II each have six chromatids that are joined as three pairs of sister chromatids. Otherwise, the steps that occur during prophase, prometaphase, metaphase, anaphase, and telophase of meiosis II are analogous to a mitotic division.

Meiosis Versus Mitosis If we compare the outcome of meiosis (see Figure 2.12) to that of mitosis (see Figure 2.8), the results are quite different. In these examples, mitosis produced two diploid daughter cells with six chromosomes each, whereas meiosis produced four haploid daughter cells with three chromosomes each. In other words, meiosis has halved the number of chromosomes per cell. **Table 2.1** compares the key differences among mitosis, meiosis I, and meiosis II.

With regard to alleles, the results of mitosis and meiosis are also different. The daughter cells produced by mitosis are genetically identical. However, the haploid cells produced by meiosis are not genetically identical to each other because they contain only one homologous chromosome from each pair. In Chapter 3, we will consider how gametes may differ in the alleles that they carry on their homologous chromosomes.

2.4 REVIEWING THE KEY CONCEPTS

- Another way for eukaryotic cells to divide is via meiosis, which produces four haploid cells. During prophase of meiosis I, homologs synapse and crossing over may occur (see Figures 2.10–2.13).

2.4 COMPREHENSION QUESTIONS

1. When does crossing over usually occur, and what is the end result?
   a. It occurs during prophase of meiosis I, and the end result is the exchange of pieces between homologous chromosomes.
   b. It occurs during prometaphase of meiosis I, and the end result is the exchange of pieces between homologous chromosomes.

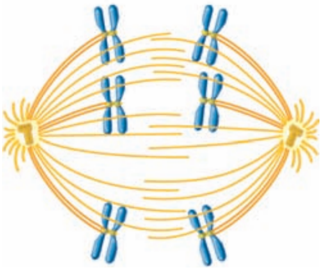
TABLE **2.1**

A Comparison of Mitosis, Meiosis I, and Meiosis II

Phase	Event	Mitosis	Meiosis I	Meiosis II
Prophase	Synapsis	No	Yes	No
Prophase	Crossing over	Rarely	Commonly	Rarely
Prometaphase	Attachment to the poles	A pair of sister chromatids to both poles	A pair of sister chromatids to one pole	A pair of sister chromatids to both poles
Metaphase	Alignment along the metaphase plate	Sister chromatids	Bivalents	Sister chromatids
Anaphase	Separation of:	Sister chromatids	Bivalents	Sister chromatids
End result		Two diploid cells		Four haploid cells

c. It occurs during prophase of meiosis I, and the end result is the separation of sister chromatids.
d. It occurs during prometaphase of meiosis I, and the end result is the separation of sister chromatids.

2. Which phase of meiosis is depicted in the drawing below?



a. Metaphase of meiosis I
b. Metaphase of meiosis II
c. Anaphase of meiosis I
d. Anaphase of meiosis II

2.5 SEXUAL REPRODUCTION

In the previous section, we considered how a diploid cell divides by meiosis to produce cells with half the genetic material of the original mother cell. This process is critical for sexual reproduction, which is a common way for eukaryotic organisms to produce offspring. During **sexual reproduction,** gametes are made that contain half the amount of genetic material. These gametes fuse with each other in the process of fertilization to begin the life of a new organism.

Gametes are highly specialized cells that are produced by a process called **gametogenesis.** As discussed previously, gametes are typically haploid, which means they contain half the number of chromosomes of diploid cells. Haploid cells are represented by 1*n* and diploid cells by 2*n*, where *n* refers to a set of chromosomes. A haploid gamete contains a single set of chromosomes, whereas a diploid cell has two sets. For example, a diploid human cell contains two sets of chromosomes, for a total of 46, but a gamete (sperm or egg cell) contains only a single set of 23 chromosomes.

Some simple eukaryotic species are **isogamous,** which means that the gametes are morphologically similar. Examples of isogamous organisms include many species of fungi and algae. Most eukaryotic species, however, are **heterogamous**—they produce two morphologically different types of gametes. Male gametes, or **sperm cells,** are relatively small and usually travel far distances to reach the female gamete—the **egg cell,** or **ovum.** The mobility of the sperm is an important characteristic, making it likely that it will come in close proximity to the egg cell. The sperm of most animal species contain a single flagellum that enables them to swim. The sperm of ferns and nonvascular plants such as bryophytes may have multiple flagella. In flowering plants, however, the sperm are contained within pollen grains. Pollen is a small mobile structure that can be carried by the wind or on the feet or hairs of insects. In flowering plants, sperm are delivered to egg cells via pollen tubes. Compared with sperm cells, the egg cell is usually very large and nonmotile. In animal species, the egg stores a large amount of nutrients to nourish the growing embryo. In this section, we will examine how sperm and egg cells are made in animal and plant species.

In Animals, Spermatogenesis Produces Four Haploid Sperm Cells and Oogenesis Produces a Single Haploid Egg Cell

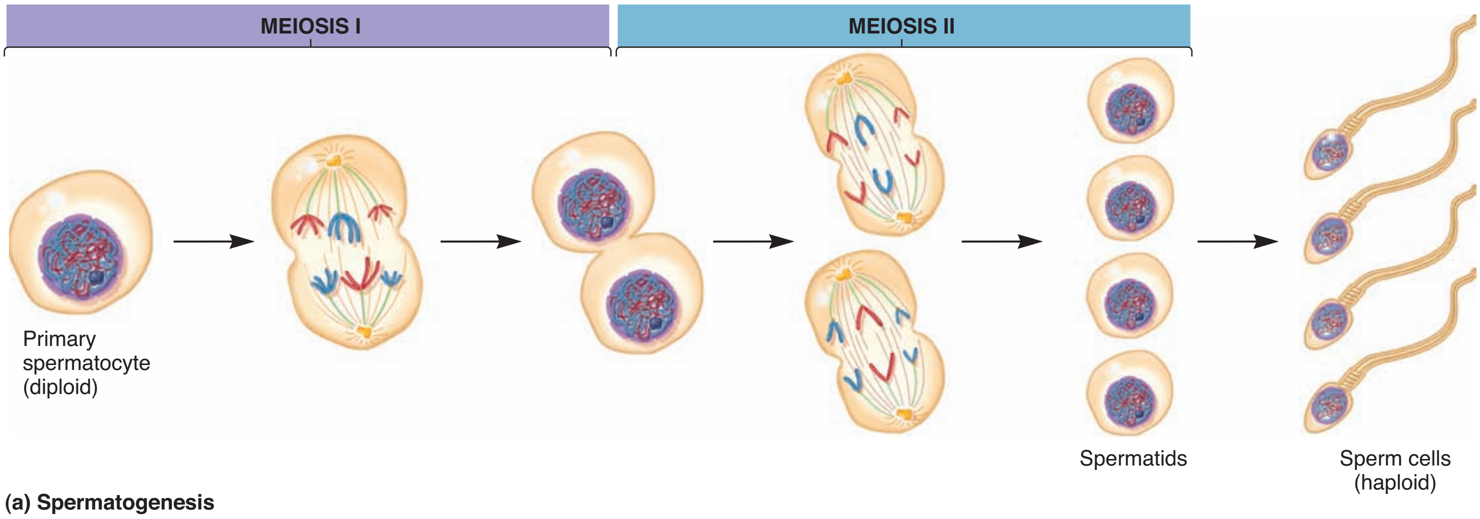
In male animals, **spermatogenesis,** the production of sperm, occurs within glands known as the testes. The testes contain spermatogonial cells that divide by mitosis to produce two cells. One of these remains a spermatogonial cell, and the other cell becomes a primary spermatocyte. As shown in **Figure 2.14a**, the spermatocyte progresses through meiosis I and meiosis II to produce four haploid cells, which are known as spermatids. These cells then mature into sperm cells. The structure of a sperm cell includes a long flagellum and a head. The head of the sperm contains little more than a haploid nucleus and an organelle at its tip, known as an acrosome. The acrosome contains digestive enzymes that are released when a sperm meets an egg cell. These enzymes enable the sperm to penetrate the outer protective layers of the egg and gain entry into the egg cell's cytosol. In animal species without a mating season, sperm production is a continuous process in mature males. A mature human male, for example, produces several hundred million sperm each day.

In female animals, **oogenesis,** the production of egg cells, occurs within specialized diploid cells of the ovary known as oogonia. Quite early in the development of the ovary, the oogonia initiate meiosis to produce primary oocytes. For example, in humans, approximately 1 million primary oocytes per ovary are produced before birth. These primary oocytes are arrested—enter a dormant phase—at prophase of meiosis I, remaining at this stage until the female becomes sexually mature. Beginning at this stage, primary oocytes are periodically activated to progress through the remaining stages of oocyte development.

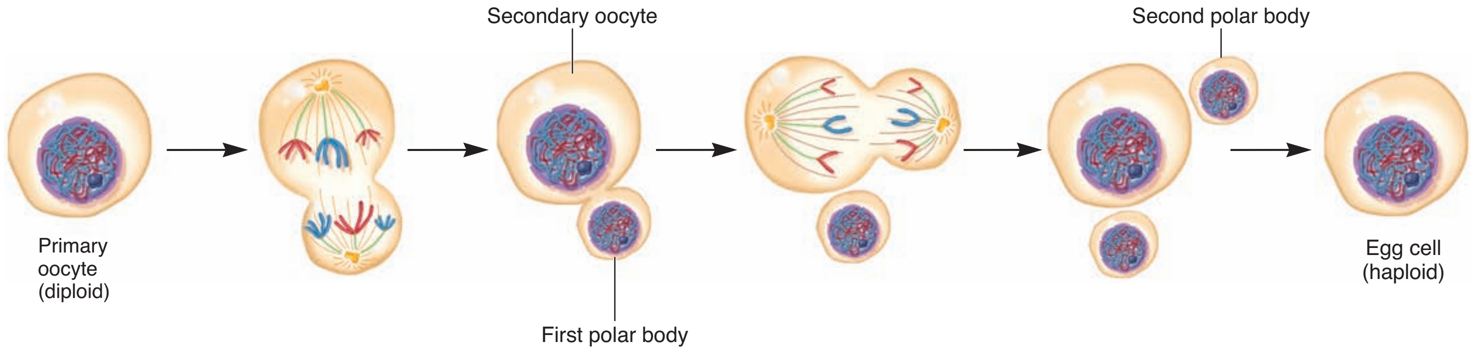
During oocyte maturation, meiosis produces only one cell that is destined to become an egg, as opposed to the four gametes produced from each primary spermatocyte during spermatogenesis. How does this occur? As shown in Figure 2.14b, the first meiotic division is asymmetrical and produces a secondary oocyte and a much smaller cell, known as a polar body. Most of the cytoplasm is retained by the secondary oocyte and very little by the polar body, allowing the oocyte to become a larger cell with more stored nutrients. The secondary oocyte then begins meiosis II. In mammals, the secondary oocyte is released from the ovary—an event called ovulation—and travels down the oviduct toward the uterus. During this journey, if a sperm cell penetrates the secondary oocyte, it is stimulated to complete meiosis II; the secondary oocyte produces a haploid egg and a second polar body. The haploid egg and sperm nuclei then unite to create the diploid nucleus of a new individual.

Plant Species Alternate Between Haploid (Gametophyte) and Diploid (Sporophyte) Generations

Most species of animals are diploid, and their haploid gametes are considered to be a specialized type of cell. By comparison, the life cycles of plant species alternate between haploid and diploid generations. The haploid generation is called the **gametophyte,** whereas the diploid generation is called the **sporophyte.** Meiosis produces haploid cells called spores, which divide by mitosis to produce the gametophyte. In simpler plants, such as mosses, a haploid spore can produce a large multicellular gametophyte by repeated mitoses



(a) Spermatogenesis



(b) Oogenesis

FIGURE 2.14 **Gametogenesis in animals. (a)** Spermatogenesis. A diploid spermatocyte undergoes meiosis to produce four haploid (n) spermatids. These differentiate during spermatogenesis to become mature sperm. **(b)** Oogenesis. A diploid oocyte undergoes meiosis to produce one haploid egg cell and two or three polar bodies. For some species, the first polar body divides; in other species, it does not. Because of asymmetrical cytokinesis, the amount of cytoplasm the egg receives is maximized. The polar bodies degenerate.

Concept check: *What are polar bodies?*

and cellular divisions. In flowering plants, however, spores develop into gametophytes that contain only a few cells. In this case, the organism that we think of as a "plant" is the sporophyte, whereas the gametophyte is very inconspicuous. In fact, the gametophytes of most plant species are small structures produced within the much larger sporophyte. Certain cells within the haploid gametophytes then become specialized as haploid gametes.

Figure 2.15 provides an overview of gametophyte development and gametogenesis in flowering plants. Meiosis occurs within two different structures of the sporophyte: the anthers and the ovaries, which produce male and female gametophytes, respectively. This diagram depicts a flower from an angiosperm, which is a plant that produces seeds within an ovary.

In the anther, diploid cells called microsporocytes undergo meiosis to produce four haploid microspores. These separate into individual microspores. In many flowering plants, each microspore undergoes mitosis to produce a two-celled structure containing one tube cell and one generative cell, both of which are haploid. This structure differentiates into a **pollen grain,** which is the male gametophyte with a thick cell wall. Later, the generative cell undergoes mitosis to produce two haploid sperm cells. In most plant species, this mitosis occurs only if the pollen grain germinates—if it lands on a stigma and forms a pollen tube (look ahead to Figure 3.2c).

By comparison, female gametophytes are produced within ovules found in the plant ovaries. A type of cell known as a megasporocyte undergoes meiosis to produce four haploid megaspores. Three of the four megaspores degenerate. The remaining haploid megaspore then undergoes three successive mitotic divisions accompanied by asymmetrical cytokinesis to produce seven individual cells—one egg cell, two synergids, three antipodals, and one central cell. This seven-celled structure, also known as the **embryo sac,** is the mature female gametophyte. Each embryo sac is contained within an ovule.

For fertilization to occur, specialized cells within the male and female gametophytes must meet. The steps of plant fertilization are described in Chapter 3. To begin this process, a pollen grain lands on a stigma (look ahead to Figure 3.2c). This stimulates the tube cell to sprout a tube that grows through the style and eventually makes contact with an ovule. As this is occurring, the generative cell undergoes mitosis to produce two haploid sperm cells. The sperm cells migrate through the pollen tube and eventually reach the ovule. One of the sperm enters the central cell, which contains the two polar nuclei. This results in a cell that is triploid ($3n$). This cell divides mitotically to produce **endosperm,** which acts as a food-storing tissue. The other sperm enters the egg cell. The egg and sperm nuclei fuse to create a diploid cell, the zygote, which becomes a plant embryo. Therefore, fertilization in flowering plants is actually a double fertilization. The result is that the endosperm, which uses a large amount of plant resources, will develop only when an egg cell has been fertilized. After fertilization is complete, the ovule develops into a seed, and the surrounding ovary develops into the fruit, which encloses one or more seeds.

When comparing animals and plants, it's interesting to consider how gametes are made. Animals produce gametes by meiosis. In contrast, plants produce gametes by mitosis. The gametophyte of plants is a haploid multicellular organism that is produced by mitotic cellular divisions of a haploid spore. Within the multicellular gametophyte, certain cells become specialized as gametes.

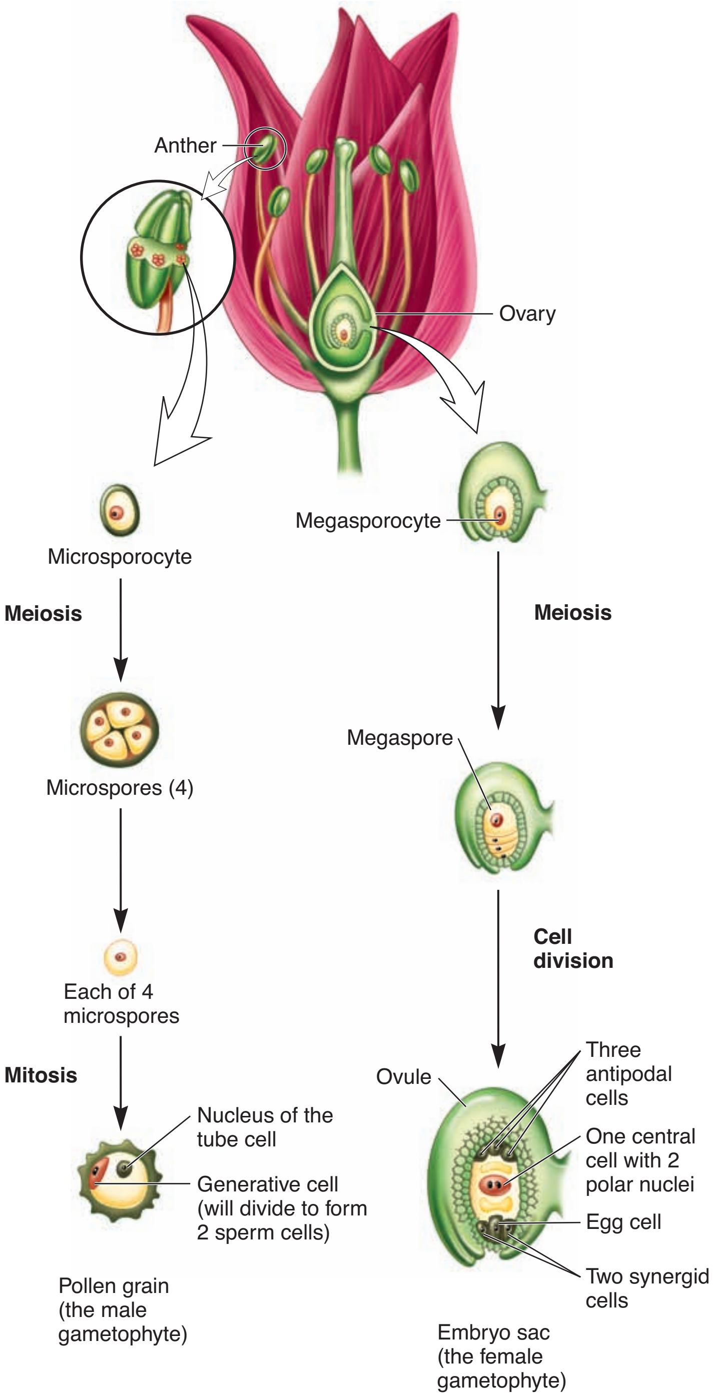


FIGURE 2.15 **The formation of male and female gametes by the gametophytes of angiosperms (flowering plants).**

Concept check: Are all of the cell nuclei in the embryo sac haploid or is just the egg haploid?

2.5 REVIEWING THE KEY CONCEPTS

- Animals produce gametes via spermatogenesis and oogenesis (see Figure 2.14).
- Plants exhibit alternation of generations between a diploid sporophyte and a haploid gametophyte. The gametophyte produces gametes (see Figure 2.15).

2.5 COMPREHENSION QUESTIONS

1. In animals, a key difference between spermatogenesis and oogenesis is
   a. only oogenesis involves meiosis.
   b. only spermatogenesis involves meiosis.
   c. spermatogenesis produces four sperm, whereas oogenesis produces only one egg cell.
   d. None of the above.
2. Which of the following statements regarding plants is false?
   a. Meiosis within anthers produces spores that develop into pollen.
   b. Meiosis within ovules produces spores that develop into an embryo sac.
   c. The male gametophyte is a pollen grain, and the female gametophyte is an embryo sac.
   d. Meiosis directly produces sperm and egg cells in plants.

KEY TERMS

Page 19. reproduction, chromosomes, chromatin
Page 20. prokaryotes, nucleoid, eukaryotes, organelles, nucleus
Page 21. cytogenetics, cytogeneticist, somatic cell, gametes, germ cells, karyotype, diploid, homologs, alleles
Page 23. homozygous, heterozygous, locus, loci, asexual reproduction
Page 24. binary fission, cell cycle, interphase, restriction point, chromatids, centromere, sister chromatids, kinetochore
Page 25. mitosis
Page 26. mitotic spindle apparatus, mitotic spindle, microtubule-organizing centers (MTOC), centrosomes, spindle pole, centrioles
Page 27. decondensed, prophase, condense, prometaphase
Page 29. metaphase plate, metaphase, anaphase, telophase, cytokinesis, cleavage furrow, cell plate
Page 30. meiosis, haploid, leptotene, zygotene, synapsis, synaptonemal complex, pachytene, bivalents, tetrad, crossing over, chiasma, chiasmata, diplotene, diakinesis
Page 34. sexual reproduction, gametogenesis, isogamous, heterogamous, sperm cells, egg cell, ovum, spermatogenesis
Page 35. oogenesis, gametophyte, sporophyte
Page 36. pollen grain, embryo sac, endosperm

CHAPTER SUMMARY

2.1 General Features of Chromosomes

- Chromosomes are structures that contain the genetic material, which is DNA.
- Prokaryotic cells are simple and lack cell compartmentalization, but eukaryotic cells contain a cell nucleus and other compartments (see Figure 2.1).
- Chromosomes can be examined under the microscope. An organized representation of the chromosomes from a single cell is called a karyotype (see Figure 2.2).
- In eukaryotic species, the chromosomes are found in sets. Eukaryotic cells are often diploid, which means that each type of chromosome occurs in a homologous pair (see Figure 2.3).

2.2 Cell Division

- Bacteria divide by binary fission (see Figure 2.4).
- To divide, eukaryotic cells progress through a cell cycle (see Figure 2.5).
- Prior to cell division, eukaryotic chromosomes are replicated to form sister chromatids (see Figure 2.6).

2.3 Mitosis and Cytokinesis

- Chromosome sorting in eukaryotes is achieved via a spindle apparatus (see Figure 2.7).
- A common way for eukaryotic cells to divide is by mitosis and cytokinesis. Mitosis is divided into prophase, prometaphase, metaphase, anaphase, and telophase (see Figures 2.8, 2.9).

2.4 Meiosis

- Another way for eukaryotic cells to divide is via meiosis, which produces four haploid cells. During prophase of meiosis I, homologs synapse and crossing over may occur (see Figures 2.10–2.13).

2.5 Sexual Reproduction

- Animals produce gametes via spermatogenesis and oogenesis (see Figure 2.14).
- Plants exhibit alternation of generations between a diploid sporophyte and a haploid gametophyte. The gametophyte produces gametes (see Figure 2.15).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe the attachments of kinetochore microtubules to sister chromatids and to the spindle poles during metaphase of mitosis, meiosis I, and meiosis II.

Answer: During mitosis and during meiosis II, each pair of sister chromatids has kinetochore microtubules that link them to both poles. By comparison, during metaphase of meiosis I, the sister chromatids line up as bivalent. Within a bivalent, kinetochore microtubules attach a pair of sister chromatids to just one pole.

S2. What are the key differences for anaphase when comparing mitosis, meiosis I, and meiosis II?

Answer: During mitosis and meiosis II, the centromeres split and individual chromatids move to their respective poles. In meiosis II the centromeres do not split. Instead, the bivalents separate and pairs of sister chromatids move to opposite poles.

S3. Among different plant species, both male and female gametophytes can be produced by single individuals or by separate sexes. In some species, such as the garden pea, a single individual can produce both male and female gametophytes. Fertilization takes place via self-fertilization or cross-fertilization. A plant species that has a single type of flower producing both pollen and eggs is termed a monoclinous plant. In other plant species, two different types of flowers produce either pollen or eggs. When both flower types are on a single individual, such a species is termed monoecious. It is most common for the "male flowers" to be produced near the top of the plant and the "female flowers" toward the bottom. Though less common, some species of plants are dioecious. For dioecious species, one individual makes either male flowers or female flowers, but not both.

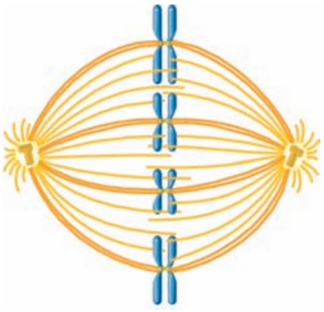
Based on your personal observations of plants, try to give examples of monoclinous, monoecious, and dioecious plants.

Answer: Monoclinous plants—pea plant, tulip, and roses. The same flower produces pollen on the anthers and egg cells within the ovary.

Monoecious plants—corn and pine trees. In corn, the tassels are the male flowers and the ears result from fertilization within the female flowers. In pine trees, pollen is produced in cones near the top of the tree, and eggs cells are found in larger cones nearer the bottom.

Dioecious plants—holly and ginkgo trees. Certain individuals produce only pollen; others produce only eggs.

S4. A diploid cell has eight chromosomes, four per set. For the following diagram, in what phase of mitosis, meiosis I or meiosis II, is this cell?



Answer: The cell is in metaphase of meiosis II. You can tell because the chromosomes are lined up in a single row along the metaphase plate, and the cell has only four pairs of sister chromatids. If it were mitosis, the cell would have eight pairs of sister chromatids.

S5. Why is it necessary for gametes to be haploid?

Answer: It is necessary for gametes to be haploid so that the chromosome number in succeeding generations stays the same. Each new member of a population receives half of its genetic material from its mother and half from its father. If the gametes contained the same number of chromosomes as somatic cells, the number of chromosomes would double with each succeeding generation.

Conceptual Questions

C1. The process of binary fission begins with a single mother cell and ends with two daughter cells. Would you expect the mother and daughter cells to be genetically identical? Explain why or why not.

C2. What is a homolog? With regard to genes and alleles, how are homologs similar to and different from each other?

C3. What is a sister chromatid? Are sister chromatids genetically similar or identical? Explain.

C4. With regard to sister chromatids, which phase of mitosis is the organization phase, and which is the separation phase?

C5. A species is diploid containing three chromosomes per set. Draw what the chromosomes would look like in the G_1 and G_2 phases of the cell cycle.

C6. How does the attachment of kinetochore microtubules to the kinetochore differ in metaphase of meiosis I from metaphase of mitosis? Discuss what you think would happen if a sister chromatid was not attached to a kinetochore microtubule.

C7. For the following events, specify whether they occur during mitosis, meiosis I, or meiosis II:

A. Separation of conjoined chromatids within a pair of sister chromatids

B. Pairing of homologous chromosomes

C. Alignment of chromatids along the metaphase plate

D. Attachment of sister chromatids to both poles

C8. Describe the key events during meiosis that result in a 50% reduction in the amount of genetic material per cell.

C9. A cell is diploid and contains three chromosomes per set. Draw the arrangement of chromosomes during metaphase of mitosis and metaphase of meiosis I and II. In your drawing, make one set dark and the other lighter.

C10. The arrangement of homologs during metaphase of meiosis I is a random process. In your own words, explain what this means.

C11. A eukaryotic cell is diploid containing 10 chromosomes (5 in each set). For mitosis and meiosis, how many daughter cells would be produced, and how many chromosomes would each one contain?

C12. If a diploid cell contains six chromosomes (i.e., three per set), how many possible random arrangements of homologs could occur during metaphase of meiosis I?

C13. A cell has four pairs of chromosomes. Assuming that crossing over does not occur, what is the probability that a gamete will contain all of the paternal chromosomes? If n equals the number of chromosomes in a set, which of the following expressions can be used to calculate the probability that a gamete will receive all of the paternal chromosomes: $(1/2)^n$, $(1/2)^{n-1}$, or $n^{1/2}$?

C14. With regard to question C13, how would the phenomenon of crossing over affect the results? In other words, would the probability of a gamete inheriting only paternal chromosomes be higher or lower? Explain your answer.

C15. Eukaryotic cells must sort their chromosomes during mitosis so that each daughter cell receives the correct number of chromosomes. Why don't bacteria need to sort their chromosomes?

C16. Why is it necessary for the chromosomes to condense during mitosis and meiosis? What do you think might happen if the chromosomes were not condensed?

C17. Nine-banded armadillos almost always give birth to four offspring that are genetically identical quadruplets. Explain how you think this happens.

C18. A diploid species contains four chromosomes per set for a total of eight chromosomes in its somatic cells. Draw the cell as it would look in late prophase of meiosis II and prophase of mitosis.

Discuss how prophase of meiosis II and prophase of mitosis differ from each other, and explain how the difference originates.

C19. Explain why the products of meiosis may not be genetically identical, whereas the products of mitosis are.

C20. The period between meiosis I and meiosis II is called interphase II. Does DNA replication take place during interphase II?

C21. List several ways in which telophase appears to be the reverse of prophase and prometaphase.

C22. Corn has 10 chromosomes per set, and the sporophyte of the species is diploid. If you performed a karyotype, what is the total number of chromosomes you would expect to see in the following types of cells?

A. A leaf cell

B. The sperm nucleus of a pollen grain

C. An endosperm cell after fertilization

D. A root cell

C23. The arctic fox has 50 chromosomes (25 per set), and the common red fox has 38 chromosomes (19 per set). These species can interbreed to produce viable but infertile offspring. How many chromosomes would the offspring have? What problems do you think may occur during meiosis that would explain the offspring's infertility?

C24. Describe the cellular differences between male and female gametes.

C25. At puberty, the testes contain a finite number of cells and produce an enormous number of sperm cells during the life span of a man. Explain why testes do not run out of spermatogonial cells.

C26. Describe the timing of meiosis I and II during human oogenesis.

C27. A phenotypically normal woman with an abnormally long chromosome 13 (and a normal homolog of chromosome 13) marries a phenotypically normal man with an abnormally short chromosome 11 (and a normal homolog of chromosome 11). What is the probability of producing an offspring that will have both a long chromosome 13 and a short chromosome 11? If such a child is produced, what is the probability that this child would eventually pass both abnormal chromosomes to one of her or his offspring?

Application and Experimental Questions

E1. When studying living cells in a laboratory, researchers sometimes use drugs as a way to make cells remain at a particular stage of the cell cycle. For example, aphidicolin is an antibiotic that inhibits DNA synthesis in eukaryotic cells and causes them to remain in the G_1 phase because they cannot replicate their DNA. In what phase of the cell cycle—G_1, S, G_2, prophase, metaphase, anaphase, or telophase—would you expect somatic cells to stay if the following types of drug were added?

A. A drug that inhibits microtubule formation

B. A drug that allows microtubules to form but prevents them from shortening

C. A drug that inhibits cytokinesis

D. A drug that prevents chromosomal condensation

E2. With regard to thickness and length, what do you think the chromosomes would look like if you microscopically examined them during interphase? How would that compare with their appearance during metaphase?

E3. A rare form of dwarfism that also includes hearing loss was found to run in a particular family. It is inherited in a dominant manner. It was discovered that an affected individual had one normal copy of chromosome 15 and one abnormal copy of chromosome 15 that was unusually long. How would you determine if the unusually long chromosome 15 was causing this disorder?

Questions for Student Discussion/Collaboration

1. A diploid eukaryotic cell has 10 chromosomes (5 per set). As a group, take turns having one student draw the cell as it would look during a phase of mitosis, meiosis I, or meiosis II; then have the other students guess which phase it is.

2. Discuss the advantages and disadvantages of asexual versus sexual reproduction.

Answers to Comprehension Questions

2.1: c, d, a
2.2: d, b, b
2.3: d, d
2.4: a, c
2.5: c, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

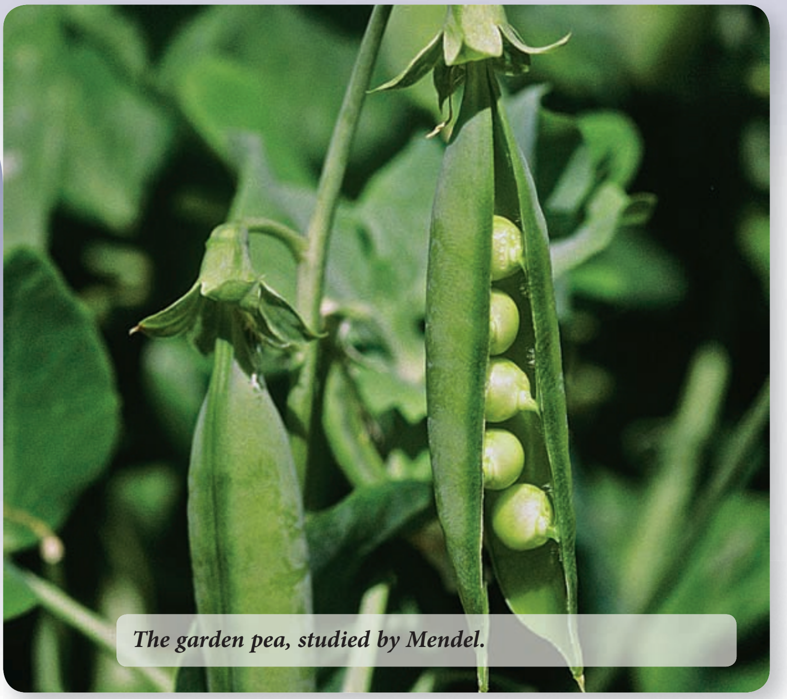
www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

3

CHAPTER OUTLINE

3.1 Mendel's Study of Pea Plants
3.2 Law of Segregation
3.3 Law of Independent Assortment
3.4 Chromosome Theory of Inheritance
3.5 Studying Inheritance Patterns in Humans
3.6 Probability and Statistics



The garden pea, studied by Mendel.

MENDELIAN INHERITANCE

An appreciation for the concept of heredity can be traced far back in human history. Hippocrates, a famous Greek physician, was the first person to provide an explanation for hereditary traits (ca. 400 B.C.E.). He suggested that "seeds" are produced by all parts of the body, which are then collected and transmitted to the offspring at the time of conception. Furthermore, he hypothesized that these seeds cause certain traits of the offspring to resemble those of the parents. This idea, known as **pangenesis,** was the first attempt to explain the transmission of hereditary traits from generation to generation.

For the next 2000 years, the ideas of Hippocrates were accepted by some and rejected by many. After the invention of the microscope in the late seventeenth century, some people observed sperm and thought they could see a tiny creature inside, which they termed a homunculus (little man). This homunculus was hypothesized to be a miniature human waiting to develop within the womb of its mother. Those who held that thought, known as spermists, suggested that only the father was responsible for creating future generations and that any resemblance between mother and offspring was due to influences "within the womb." During the same time, an opposite school of thought also developed. According to the ovists, the egg was solely responsible for human characteristics. The only role of the sperm was to stimulate the egg onto its path of development. Of course, neither of these ideas was correct.

The first systematic studies of genetic crosses were carried out by Joseph Kölreuter from 1761 to 1766. In crosses between different strains of tobacco plants, he found that the offspring were usually intermediate in appearance between the two parents. This led Kölreuter to conclude that both parents make equal genetic contributions to their offspring. Furthermore, his observations were consistent with the **blending hypothesis of inheritance.** According to this view, the factors that dictate hereditary traits can blend together from generation to generation. The blended traits are then passed to the next generation. The popular view before the 1860s, which combined the notions of pangenesis and blending inheritance, was that hereditary traits were rather malleable and could change and blend over the course of one or two generations. However, the pioneering work of Gregor Mendel would prove instrumental in refuting this viewpoint.

In Chapter 3, we will first examine the outcome of Mendel's crosses in pea plants. We begin our inquiry into genetics here because the inheritance patterns observed in peas are fundamentally related to inheritance patterns found in other eukaryotic species, such as corn, fruit flies, mice, and humans. We will discover how Mendel's insights into the patterns of inheritance in pea plants revealed some simple rules that govern the process of inheritance. In Chapters 4 through 9, we will explore more complex patterns of inheritance and also consider the role that chromosomes play as the carriers of the genetic material.

In the last part of this chapter, we will become familiar with general concepts in probability and statistics. How are statistical methods useful? Probability calculations allow us to predict the outcomes of simple genetic crosses, as well as the outcomes of more complicated crosses described in later chapters. In addition, we will learn how to use statistics to test the validity of genetic hypotheses that attempt to explain the inheritance patterns of traits.

3.1 MENDEL'S STUDY OF PEA PLANTS

Gregor Johann Mendel, born in 1822, is now remembered as the father of genetics (**Figure 3.1**). He grew up on a farm in Hynčice (formerly Heinzendorf) in northern Moravia, which was then a part of Austria and is now a part of the Czech Republic. As a young boy, he worked with his father grafting trees to improve the family orchard. Undoubtedly, his success at grafting taught him that precision and attention to detail are important elements of success. These qualities would later be important in his experiments as an adult scientist. Instead of devoting himself to farming, however, Mendel was accepted into the Augustinian monastery of St. Thomas, completed his studies for the priesthood, and was ordained in 1847. Soon after becoming a priest, Mendel worked for a short time as a substitute teacher. To continue that role, he needed to obtain a teaching license from the government. Surprisingly, he failed the licensing exam due to poor answers in the areas of physics and natural history. Therefore, Mendel then enrolled at the University of Vienna to expand his knowledge in these two areas. Mendel's training in physics and mathematics taught him to perceive the world as an orderly place, governed by natural laws. In his studies, Mendel learned that these natural laws could be stated as simple mathematical relationships.

In 1856, Mendel began his historic studies on pea plants. For 8 years, he grew and crossed thousands of pea plants on a small 23- by 115-foot plot. He kept meticulously accurate records that included quantitative data concerning the outcome of his crosses. He published his work, entitled *Experiments on Plant Hybrids,* in 1866. This paper was largely ignored by scientists at that time, possibly because of its title, which gave little indication of its contents. Another reason his work went unrecognized could be tied to a lack of understanding of chromosomes and their transmission, a topic we discussed in Chapter 2. Nevertheless, Mendel's groundbreaking work allowed him to propose the natural laws that now provide a framework for our understanding of genetics.

FIGURE 3.1 **Gregor Johann Mendel, the father of genetics.**

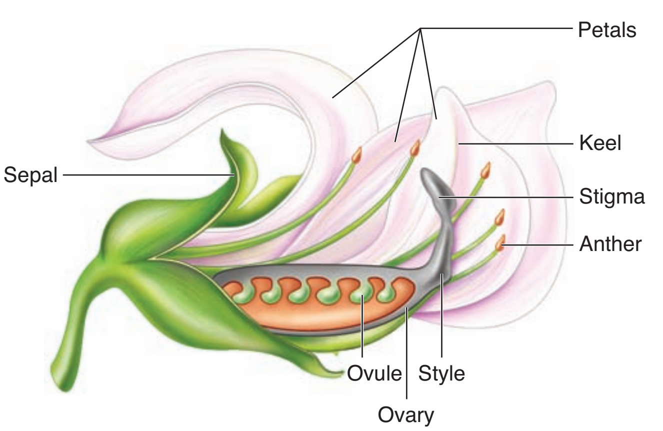
Prior to his death in 1884, Mendel reflected, "My scientific work has brought me a great deal of satisfaction and I am convinced that it will be appreciated before long by the whole world." Sixteen years later, in 1900, the work of Mendel was independently rediscovered by three biologists with an interest in plant genetics: Hugo de Vries of Holland, Carl Correns of Germany, and Erich von Tschermak of Austria. Within a few years, the influence of Mendel's studies was felt around the world. In this section, we will examine Mendel's experiments and consider their monumental significance in the field of genetics.

Mendel Chose Pea Plants as His Experimental Organism

Mendel's study of genetics grew out of his interest in ornamental flowers. Prior to his work with pea plants, many plant breeders had conducted experiments aimed at obtaining flowers with new varieties of colors. When two distinct individuals with different characteristics are mated, or **crossed,** to each other—a process called a **hybridization** experiment—their offspring are referred to as **hybrids.** For example, a hybridization experiment could involve a cross between a purple-flowered plant and a white-flowered plant. Mendel was particularly intrigued, in such experiments, by the consistency with which offspring of subsequent generations showed characteristics of one or the other parent. His intellectual foundation in physics and the natural sciences led him to consider that this regularity might be rooted in natural laws that could be expressed mathematically. To uncover these laws, he realized that he needed to carry out quantitative experiments in which the numbers of offspring carrying certain traits were carefully recorded and analyzed.

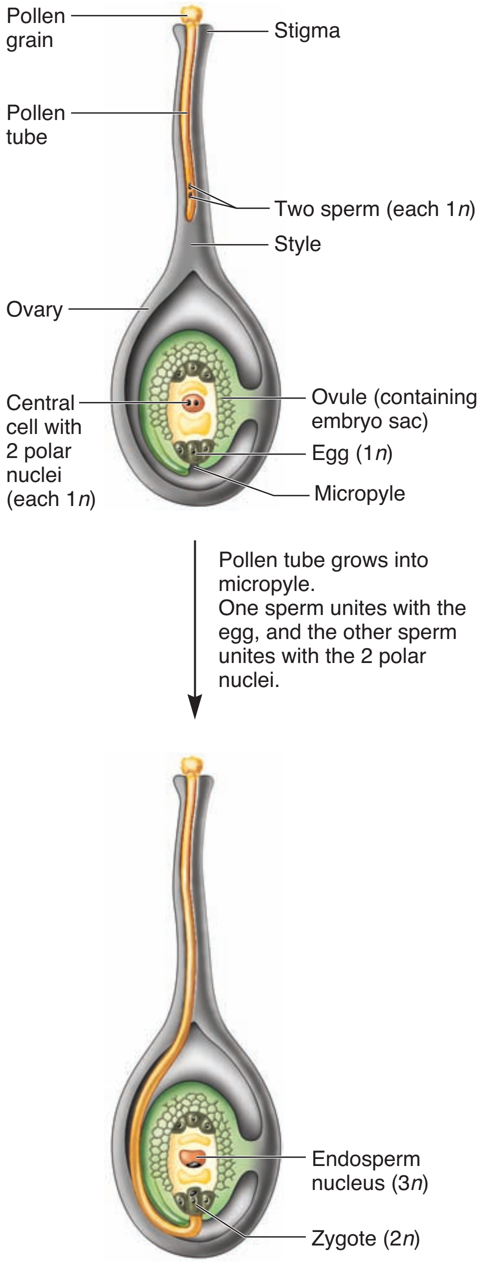
Mendel chose the garden pea, *Pisum sativum*, to investigate the natural laws that govern plant hybrids. The morphological features of this plant are shown in **Figure 3.2a, b**. Several properties of this species were particularly advantageous for studying plant hybridization. First, the species was available in several varieties that had decisively different physical characteristics. Many strains of the garden pea were available that varied in height and in the appearance of their flowers, seeds, and pods.

A second important issue is the ease of making crosses. In flowering plants, sexual reproduction occurs by a pollination event (**Figure 3.2c**). The term **gamete** is used to describe haploid reproductive cells that can unite to form a zygote. In plants, male gametes **(sperm)** are produced within **pollen grains** formed in the **anthers,** and the female gametes **(eggs)** are contained within **ovules** that form in the **ovaries.** For fertilization to occur, a pollen grain lands on the **stigma,** which stimulates the growth of a pollen


Petals
Keel
Sepal
Stigma
Anther
Ovule
Style
Ovary

(a) Structure of a pea flower

(b) A flowering pea plant


Pollen grain
Stigma
Pollen tube
Two sperm (each 1n)
Style
Ovary
Central cell with 2 polar nuclei (each 1n)
Ovule (containing embryo sac)
Egg (1n)
Micropyle
Pollen tube grows into micropyle.
One sperm unites with the egg, and the other sperm unites with the 2 polar nuclei.
Endosperm nucleus (3n)
Zygote (2n)

(c) Pollination and fertilization in angiosperms

FIGURE 3.2 **Flower structure and pollination in pea plants. (a)** The pea flower can produce both pollen and egg cells. The pollen grains are produced within the anthers, and the egg cells are produced within the ovules that are contained within the ovary. A modified petal called a keel encloses the anthers and ovaries. **(b)** Photograph of a flowering pea plant. **(c)** A pollen grain must first land on the stigma. After this occurs, the pollen sends out a long tube through which two sperm cells travel toward an ovule to reach an egg cell. The fusion between a sperm and an egg cell results in fertilization and creates a zygote. A second sperm fuses with a central cell containing two polar nuclei to create the endosperm. The endosperm provides a nutritive material for the developing embryo.

Concept check: Prior to fertilization, where is the male gamete located?

tube. This enables sperm cells to enter the stigma and migrate toward an ovule. Fertilization occurs when a sperm enters the micropyle, an opening in the ovule wall, and fuses with an egg cell.

In some experiments, Mendel wanted to carry out **self-fertilization,** which means that the pollen and egg are derived from the same plant. In peas, a modified petal known as the keel covers the reproductive structures of the plant. Because of this covering, pea plants naturally reproduce by self-fertilization. Usually, pollination occurs even before the flower opens. In other experiments, however, Mendel wanted to make crosses between different plants. How did he accomplish this goal? Fortunately, pea plants contain relatively large flowers that are easy to manipulate, making it possible to make crosses between two particular plants and study their outcomes. This process, known as **cross-fertilization** (shown in Figure 3.3), requires that the pollen from one plant be placed on the stigma of another plant. Mendel was able to pry open immature flowers and remove the anthers before they produced pollen. Therefore, these flowers could not self-fertilize. He then obtained pollen from another plant by gently touching its mature anthers with a paintbrush. Mendel applied this pollen to the stigma of the flower that already had its anthers removed. In this way, he was able to cross-fertilize his pea plants and thereby obtain any type of hybrid he wanted.

Mendel Studied Seven Characters That Bred True

When he initiated his studies, Mendel obtained several varieties of peas that were considered to be distinct. These plants were different with regard to many morphological characteristics. Such general characteristics of an organism are called **characters.** The terms **trait,** or **variant,** are typically used to describe the specific properties of a character. For example, eye color is a character of humans and blue eyes is a trait (or variant) found in some people. Over the course of 2 years, Mendel tested his strains to determine if their characters bred true. This means that a trait did not vary in appearance from generation to generation. For example, if the seeds from a pea plant were yellow, the next generation would also produce yellow seeds. Likewise, if these offspring were allowed to self-fertilize, all of their offspring would also produce yellow seeds, and so on. A variety that continues to produce the same characteristic after several generations of self-fertilization is called a **true-breeding line,** or **strain.**

Mendel next concentrated his efforts on the analysis of characteristics that were clearly distinguishable between different true-breeding lines. Figure 3.4 illustrates the seven characters that Mendel eventually chose to follow in his breeding experiments. All seven were found in two variants. For example, one characteristic he followed was height, which was found in two variants: tall and dwarf plants.

3.1 REVIEWING THE KEY CONCEPTS

- Mendel chose pea plants as his experimental organism (see Figures 3.1, 3.2).
- Using pea plants, it was easy for Mendel to carry out self-fertilization or cross-fertilization experiments (see Figure 3.3).
- Another advantage of using pea plants is they were available in several varieties in which a character existed in two distinct variants (see Figure 3.4).

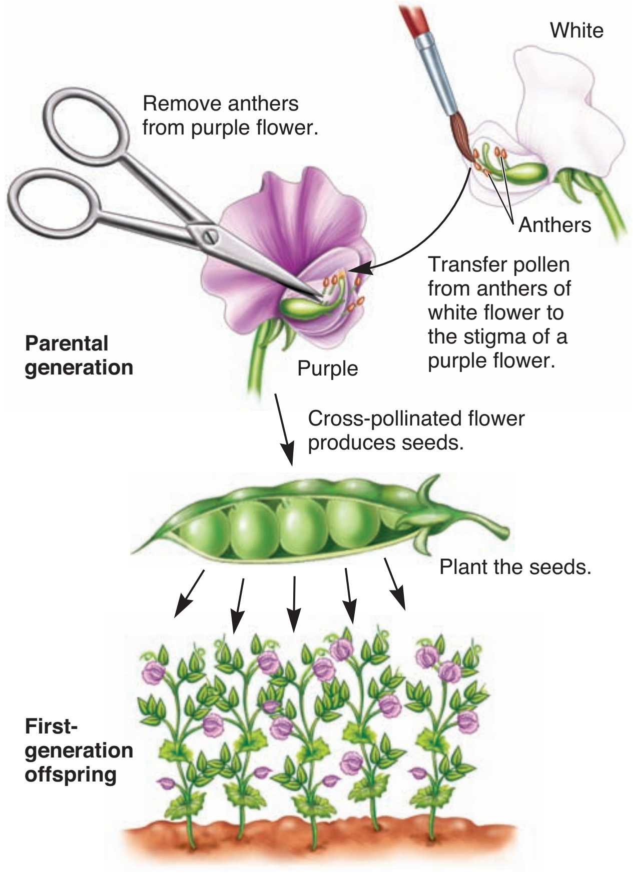

White
Remove anthers from purple flower.
Anthers
Transfer pollen from anthers of white flower to the stigma of a purple flower.
Parental generation
Purple
Cross-pollinated flower produces seeds.
Plant the seeds.
First-generation offspring

FIGURE 3.3 How Mendel cross-fertilized two different pea plants. This illustration depicts a cross between a plant with purple flowers and another plant with white flowers. The offspring from this cross are the result of pollination of the purple flower using pollen from a white flower.

Concept check: In this experiment, which plant, the white- or purple-flowered plant, is providing the egg cells, and which is providing the sperm cells?

3.1 COMPREHENSION QUESTIONS

1. Experimental advantages of using pea plants include which of the following?
   a. They came in several different varieties.
   b. They were capable of self-fertilization.
   c. They were easy to cross.
   d. All of the above were advantages.
2. The term *cross* refers to an experiment in which
   a. the gametes come from different individuals.
   b. the gametes come from a single flower of the same individual.
   c. the gametes come from different flowers of the same individual.
   d. Both a and c.
3. To avoid self-fertilization in his pea plants, Mendel had to
   a. spray the plants with a chemical that damaged the pollen.
   b. remove the anthers from immature flowers.
   c. grow the plants in a greenhouse that did not contain pollinators (e.g., bees).
   d. Do all of the above.

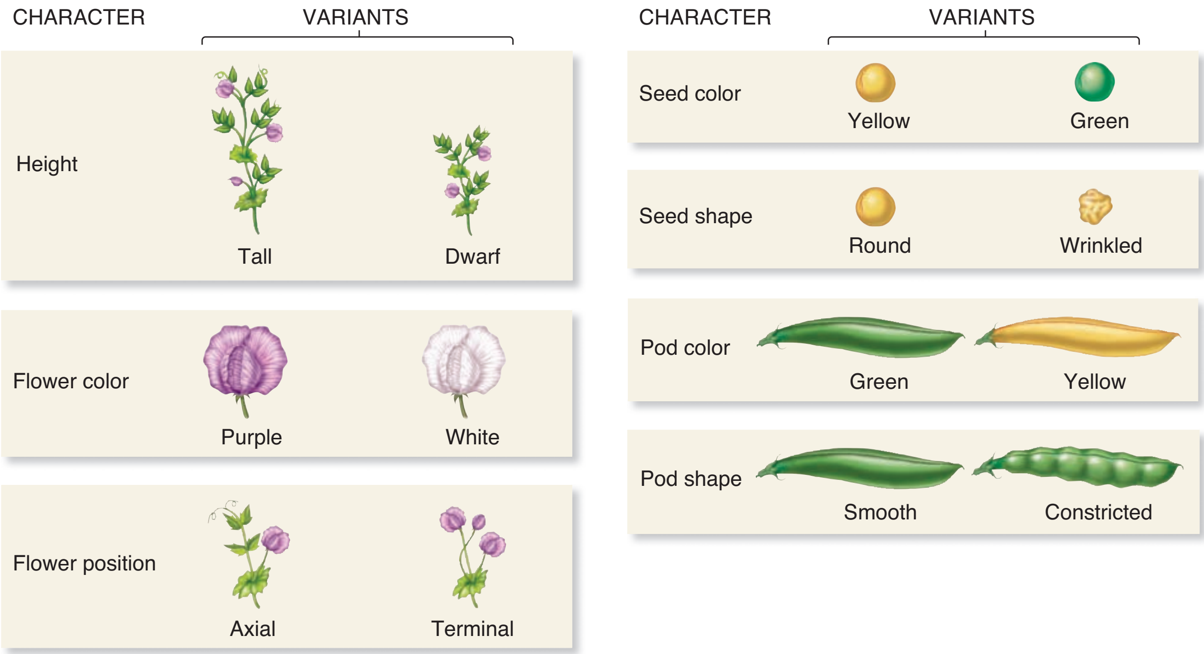


FIGURE 3.4 **An illustration of the seven characters that Mendel studied.** Each character was found as two variants that were decisively different from each other.

Concept check: *What do we mean when we say a strain is true-breeding?*

3.2 LAW OF SEGREGATION

As discussed in the previous section, Mendel was able to carry out self-fertilization or cross-fertilization experiments with his pea plants. In this section, we will examine how he studied the inheritance of characters by crossing the variants to each other. A cross in which an experimenter is observing only one character is called a **single-factor cross.** When the two parents are different variants for a given character, this type of cross produces single-character hybrids, also known as **monohybrids.** As you will learn, this type of experimental approach led Mendel to propose the law of segregation.

Mendel Followed the Outcome of a Single Character for Two Generations

Prior to conducting his studies, Mendel did not already have a hypothesis to explain the formation of hybrids. However, his educational background led him to realize that a quantitative analysis of crosses might uncover mathematical relationships that would otherwise be mysterious. His experiments were designed to determine the relationships that govern hereditary traits. This rationale is called an **empirical approach.** Laws that are deduced from an empirical approach are known as **empirical laws.**

Mendel's experimental procedure is shown in Figure 3.5. He began with true-breeding plants that differed with regard to a single character. These plants are termed the **parental generation,** or **P generation.** The true-breeding parents were crossed to each other—a process called a P cross—producing offspring called the **F_1 generation,** for first filial generation. As seen in the data, all plants of the F_1 generation showed the phenotype of one parent but not the other. This prompted Mendel to follow the transmission of this character for one additional generation. To do so, the plants of the F_1 generation were allowed to self-fertilize to produce a second generation, called the **F_2 generation,** for second filial generation.

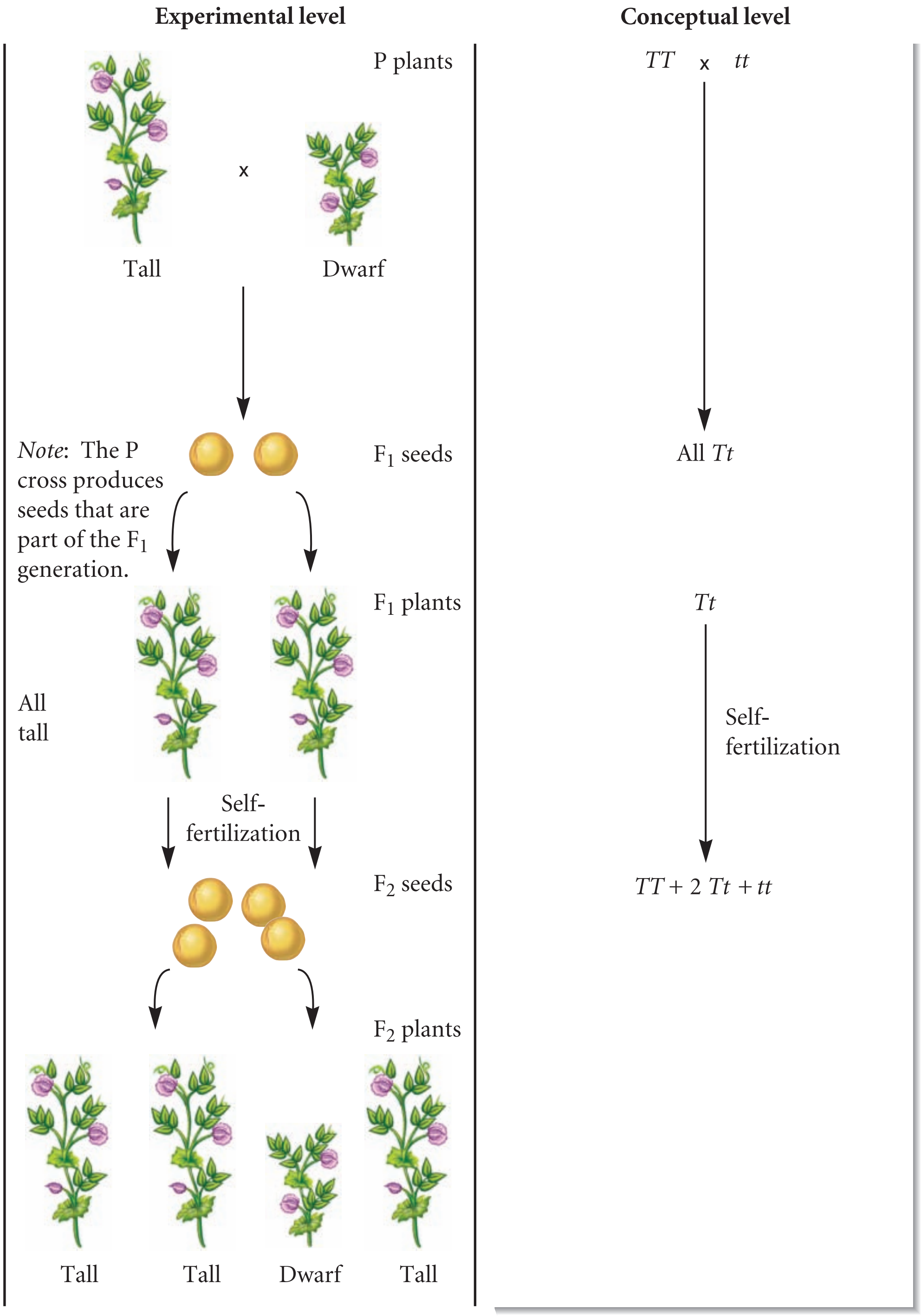
▶ THE GOAL

Mendel speculated that the inheritance pattern for a single character may follow quantitative natural laws. The goal of this experiment was to uncover such laws.

▶ ACHIEVING THE GOAL — FIGURE 3.5 Mendel's analysis of single-factor crosses.

Starting material: Mendel began his experiments with true-breeding pea plants that varied with regard to only one of seven different characters (see Figure 3.4).

1. For each of seven characters, Mendel cross-fertilized two different true-breeding lines. Keep in mind that each cross involved two plants that differed in regard to only one of the seven characters studied. The illustration at the right shows one cross between a tall and dwarf plant. This is called a P (parental) cross.
2. Collect many seeds. The following spring, plant the seeds and allow the plants to grow. These are the plants of the F_1 generation.
3. Allow the F_1 generation plants to self-fertilize. This produces seeds that are part of the F_2 generation.
4. Collect the seeds and plant them the following spring to obtain the F_2 generation plants.
5. Analyze the characteristics found in each generation.



Experimental level
Conceptual level
P plants
Tall
x
Dwarf
TT x tt
Note: The P cross produces seeds that are part of the F1 generation.
F1 seeds
All Tt
F1 plants
All tall
Tt
Self-fertilization
Self-fertilization
F2 seeds
TT + 2 Tt + tt
F2 plants
Tall
Tall
Dwarf
Tall

▶ THE DATA

P cross	*F_1 generation*	*F_2 generation*	*Ratio*
Tall × dwarf stem	All tall	787 tall, 277 dwarf	2.84:1
Purple × white flowers	All purple	705 purple, 224 white	3.15:1
Axial × terminal flowers	All axial	651 axial, 207 terminal	3.14:1
Yellow × green seeds	All yellow	6022 yellow, 2001 green	3.01:1
Round × wrinkled seeds	All round	5474 round, 1850 wrinkled	2.96:1
Green × yellow pods	All green	428 green, 152 yellow	2.82:1
Smooth × constricted pods	All smooth	882 smooth, 299 constricted	2.95:1
Total	All dominant	14,949 dominant, 5010 recessive	2.98:1

Data from: Mendel, Gregor. 1866 Versüche uber Plflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn, Bd IV für das Jahr 1865*, Abhandlungen, 3–47.

▶ INTERPRETING THE DATA

The data shown in Figure 3.5 are the results of producing an F_1 generation via cross-fertilization and an F_2 generation via self-fertilization of the F_1 monohybrids. A quantitative analysis of these data allowed Mendel to propose three important ideas:

1. Mendel's data argued strongly against a blending mechanism of inheritance. In all seven cases, the F_1 generation displayed characteristics that were distinctly like one of the two parents rather than traits intermediate in character. His first proposal was that the variant for one character is **dominant** to another variant. For example, the variant of green pods is dominant to that of yellow pods. The term **recessive** is used to describe a variant that is masked by the presence of a dominant trait but reappears in subsequent generations. Yellow pods and dwarf stems are examples of recessive variants. They can also be referred to as recessive traits.
2. When a true-breeding plant with a dominant trait was crossed to a true-breeding plant with a recessive trait, the dominant trait was always observed in the F_1 generation. In the F_2 generation, some offspring displayed the dominant trait, but a smaller proportion showed the recessive trait. How did Mendel explain this observation? Because the recessive trait appeared in the F_2 generation, he made a second proposal—the genetic determinants of traits are passed along as "unit factors" from generation to generation. His data were consistent with a **particulate theory of inheritance,** in which the genes that govern traits are inherited as discrete units that remain unchanged as they are passed from parent to offspring. Mendel called them unit factors, but we now call them genes.
3. When Mendel compared the numbers of dominant and recessive traits in the F_2 generation, he noticed a recurring pattern. Within experimental variation, he always observed approximately a 3:1 ratio between the dominant trait and the recessive trait. Mendel was the first scientist to apply this type of quantitative analysis in a biological experiment. As described next, this quantitative approach allowed him to make a third proposal—genes **segregate** from each other during the process that gives rise to gametes.

Mendel's 3:1 Phenotypic Ratio Is Consistent with the Law of Segregation

Mendel's research was aimed at understanding the laws that govern the inheritance of traits. At that time, scientists did not understand the molecular composition of the genetic material or its mode of transmission during gamete formation and fertilization. We now know that the genetic material is composed of deoxyribonucleic acid (DNA), a component of chromosomes. Each chromosome contains hundreds or thousands of shorter segments that function as genes—a term that was originally coined by the Danish botanist Wilhelm Johannsen in 1909. A **gene** is defined as a "unit of heredity" that may influence the outcome of an organism's traits. Each of the seven characters that Mendel studied is influenced by a different gene.

As discussed in Chapter 2, most eukaryotic species, such as pea plants and humans, have their genetic material organized into pairs of chromosomes. For this reason, eukaryotes have two copies of most genes. These copies may be the same or they may differ. The term **allele** refers to different versions of the same gene. With this modern knowledge, the results shown in Figure 3.5 are consistent with the idea that each parent transmits only one copy of each gene (i.e., one allele) to each offspring. **Mendel's law of segregation** states:

> *The two copies of a gene segregate (or separate) from each other during transmission from parent to offspring.*

Therefore, only one copy of each gene is found in a gamete. At fertilization, two gametes combine randomly, potentially producing different allelic combinations.

Let's use Mendel's cross of tall and dwarf pea plants to illustrate how alleles are passed from parents to offspring (**Figure 3.6**). The letters *T* and *t* are used to represent the alleles of the gene that determines plant height. By convention, the uppercase letter represents the dominant allele (*T* for tall height, in this case), and the recessive allele is represented by the same letter in lowercase (*t*, for dwarf height). For the P cross, both parents are true-breeding plants. Therefore, we know each has identical copies of the height gene. When an individual possesses two identical copies of a gene, the individual is said to be **homozygous** with respect to that gene. (The prefix *homo-* means "like," and the suffix *-zygo* means "pair.") In the P cross, the tall plant is homozygous for the tall allele *T*, and the dwarf plant is homozygous for the dwarf allele *t*. The term **genotype** refers to the genetic composition of an individual. *TT* and *tt* are the genotypes of the P generation in this experiment. The term **phenotype** refers to an observable characteristic of an organism. In the P generation, the plants exhibit a phenotype that is either tall or dwarf.

In contrast, the F_1 generation is **heterozygous,** with the genotype *Tt*, because every individual carries one copy of the tall

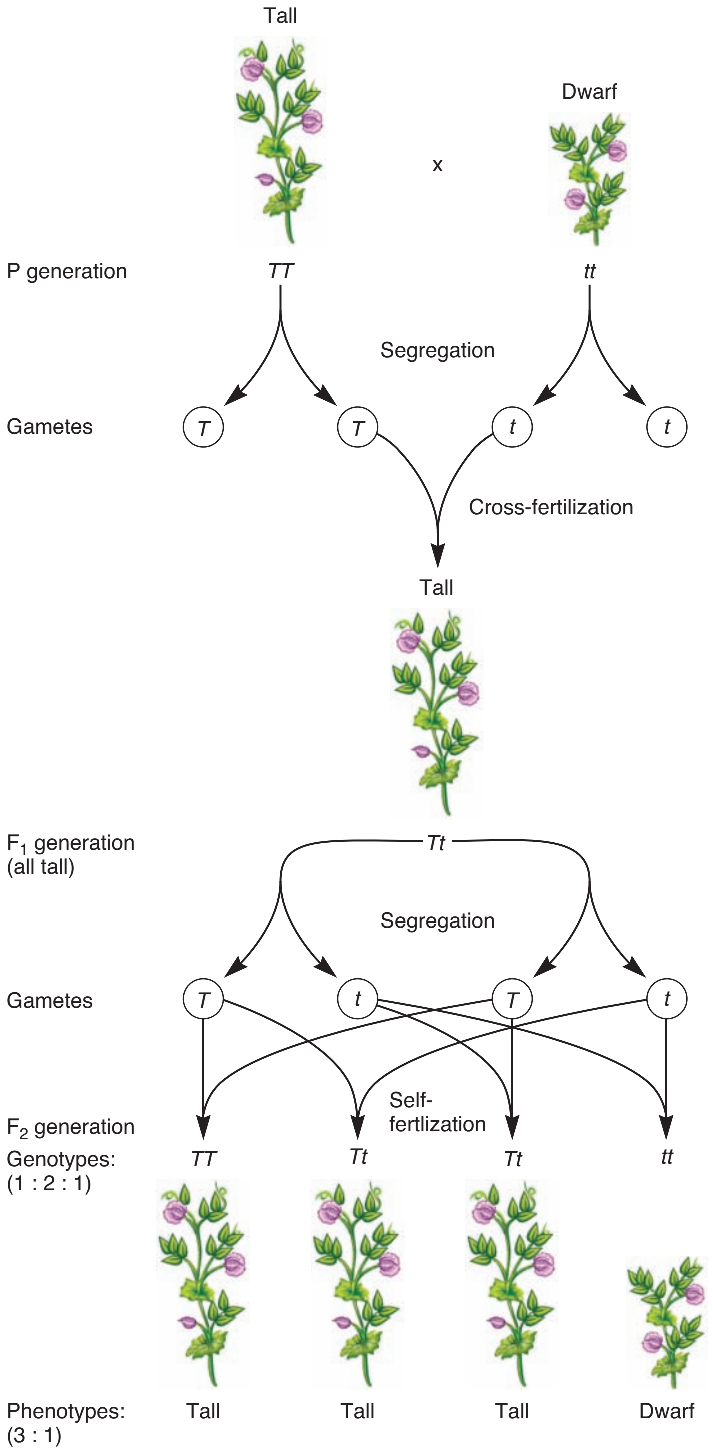


FIGURE 3.6 **Mendel's law of segregation.** This illustration shows a cross between a true-breeding tall plant and a true-breeding dwarf plant and the subsequent segregation of the tall (*T*) and dwarf (*t*) alleles in the F_1 and F_2 generations.

Concept check: With regard to the T and t alleles, explain what the word segregation means.

allele and one copy of the dwarf allele. A heterozygous individual carries different alleles of a gene. (The prefix *hetero-* means "different.") Although these plants are heterozygous, their phenotypes are tall because they have a copy of the dominant tall allele.

The law of segregation predicts that the phenotypes of the F_2 generation will be tall and dwarf in a ratio of 3:1 (see Figure 3.6). Because the parents of the F_2 generation are heterozygous, each gamete can transmit either a *T* allele or a *t* allele to a particular offspring, but not both. Therefore, *TT*, *Tt*, and *tt* are the possible genotypes of the F_2 generation. By randomly combining these alleles, the genotypes are produced in a 1:2:1 ratio. Because *TT* and *Tt* both produce tall phenotypes, a 3:1 phenotypic ratio is observed in the F_2 generation.

A Punnett Square Can Be Used to Predict the Outcome of Crosses

An easy way to predict the outcome of simple genetic crosses is to use a **Punnett square,** a method originally proposed by Reginald Punnett. To construct a Punnett square, you must know the genotypes of the parents. With this information, the Punnett square enables you to predict the types of offspring the parents are expected to produce and in what proportions. The following is a step-by-step description of the Punnett square approach using a cross of heterozygous tall plants as an example.

Step 1. *Write down the genotypes of both parents.* In this example, a heterozygous tall plant is crossed to another heterozygous tall plant. The plant providing the pollen is considered the male parent and the plant providing the eggs, the female parent.

Male parent: *Tt*

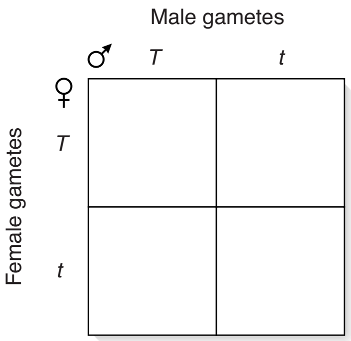
Female parent: *Tt*

Step 2. *Write down the possible gametes that each parent can make.* Remember that the law of segregation tells us that a gamete can carry only one copy of each gene.

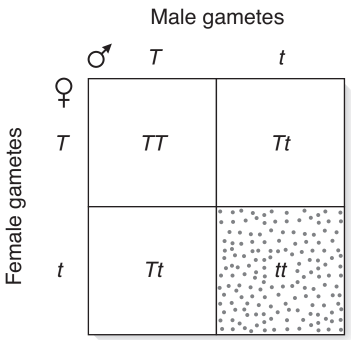
Male gametes: *T* or *t*

Female gametes: *T* or *t*

Step 3. *Create an empty Punnett square.* In the examples shown in this textbook, the number of columns equals the number of male gametes, and the number of rows equals the number of female gametes. Our example has two rows and two columns. Place the male gametes across the top of the Punnett square and the female gametes along the side.



Step 4. *Fill in the possible genotypes of the offspring by combining the alleles of the gametes in the empty boxes.*



Step 5. *Determine the relative proportions of genotypes and phenotypes of the offspring.* The genotypes are obtained directly from the Punnett square. They are contained within the boxes that have been filled in. In this example, the genotypes are *TT*, *Tt*, and *tt* in a 1:2:1 ratio. To determine the phenotypes, you must know the dominant/recessive relationship between the alleles. For plant height, we know that *T* (tall) is dominant to *t* (dwarf). The genotypes *TT* and *Tt* are tall, whereas the genotype *tt* is dwarf. Therefore, our Punnett square shows us that the ratio of phenotypes is 3:1, or 3 tall plants : 1 dwarf plant.

3.2 REVIEWING THE KEY CONCEPTS

- In the experiments described in this section, Mendel conducted single-factor crosses in which he followed the variants for a single character (see Figure 3.5).
- The results of his single-factor crosses showed that the dominant trait was always observed in the F_1 generation and displayed a 3:1 ratio in the F_2 generation.
- Based on the results of his single-factor crosses, Mendel proposed three key ideas regarding inheritance. (1) Traits may be dominant or recessive. (2) Genes are passed unaltered from generation to generation. (3) The two copies of a given gene segregate (or separate) from each other during transmission from parent to offspring.
- This third idea is known as the law of segregation (see Figure 3.6).
- A Punnett square can be used to deduce the outcome of monohybrid crosses.

3.2 COMPREHENSION QUESTIONS

1. A pea plant is *Tt*. Which of the following statements is correct?
   a. Its genotype is *Tt,* and its phenotype is dwarf.
   b. Its phenotype is *Tt,* and its genotype is dwarf.
   c. Its genotype is *Tt,* and its phenotype is tall.
   d. Its phenotype is *Tt,* and genotype is tall.
2. A *Tt* plant is crossed to a *tt* plant. What is the expected outcome for the ratio of offspring from this cross?
   a. 3 tall : 1 dwarf
   b. 1 tall : 1 dwarf
   c. 1 tall : 3 dwarf
   d. 2 tall : 1 dwarf
3. You may need to refer back to Chapter 2 to answer this question. If a plant is *Tt*, at which stage do the *T* and *t* alleles segregate from each other?
   a. Anaphase of mitosis
   b. Anaphase of meiosis I
   c. Anaphase of meiosis II
   d. None of the above.

3.3 LAW OF INDEPENDENT ASSORTMENT

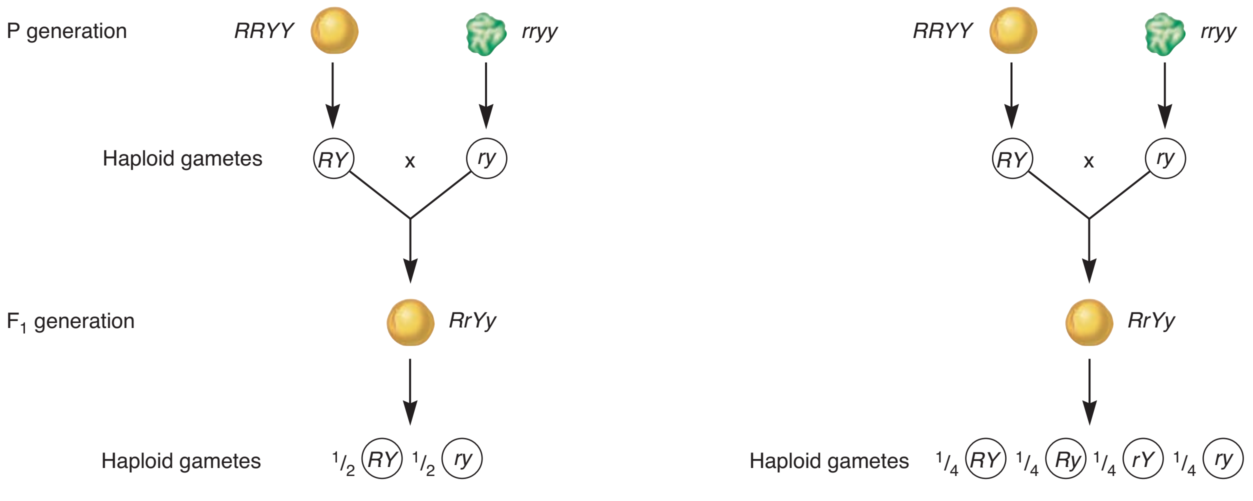
Though his experiments described in Figure 3.5 revealed important ideas regarding hereditary laws, Mendel realized that additional insights might be uncovered if he conducted more complicated experiments. In this section, we will examine how he conducted crosses in which he simultaneously investigated the pattern of inheritance for two different characters. In other words, he carried out **two-factor crosses** in which he followed the inheritance of two different characters within the same groups of individuals. These experiments led to the formulation of a second law called the law of independent assortment.

Mendel Also Analyzed Crosses Involving Two Different Characters

Let's consider one of Mendel's experiments in which one of the characters was seed shape, found in round or wrinkled variants; the second character was seed color, which existed as yellow and green variants. In this two-factor cross, Mendel followed the inheritance pattern for both characters simultaneously.

What results are possible from a two-factor cross? One possibility is that the genetic determinants for two different characters are always linked to each other and inherited as a single unit (**Figure 3.7a**). If this were the case, the F_1 offspring could produce only two types of gametes: *RY* and *ry.* A second possibility is that they are not linked and can assort themselves independently into haploid gametes (**Figure 3.7b**). According to the independent-assortment hypothesis, an F_1 offspring could produce four types of gametes: *RY*, *Ry*, *rY*, and *ry*. Keep in mind that the results of Figure 3.5 have already shown us that a gamete carries only one allele for each gene.

The experimental protocol for one of Mendel's two-factor crosses is shown in **Figure 3.8**. He began with two different strains of true-breeding pea plants that were different with regard to two characters: seed shape and seed color. In this example, one plant was produced from seeds that were round and yellow; the other plant from seeds that were wrinkled and green. When these plants were crossed, the seeds, which contain the plant embryo, are considered part of the F_1 generation. As


P generation
RRYY
rryy
Haploid gametes
RY
x
ry
F1 generation
RrYy
Haploid gametes
1/2 RY 1/2 ry
RRYY
rryy
RY
x
ry
RrYy
Haploid gametes
1/4 RY 1/4 Ry 1/4 rY 1/4 ry

(a) HYPOTHESIS: Linked assortment

(b) HYPOTHESIS: Independent assortment

FIGURE 3.7 **Two hypotheses to explain how two different genes assort during gamete formation. (a)** According to the linked hypothesis, the two genes always stay associated with each other. **(b)** In contrast, the independent-assortment hypothesis proposes that the two different genes randomly segregate into haploid cells.

Concept check: According to the linkage hypothesis shown here, what is linked? Are two different genes linked or are two different alleles of the same gene linked, or both?

expected, the data revealed that all F_1 seeds displayed a phenotype of round and yellow. This was observed because round and yellow are dominant traits. It is the F_2 generation that supports the independent-assortment model and refutes the linkage model.

▶ THE HYPOTHESES

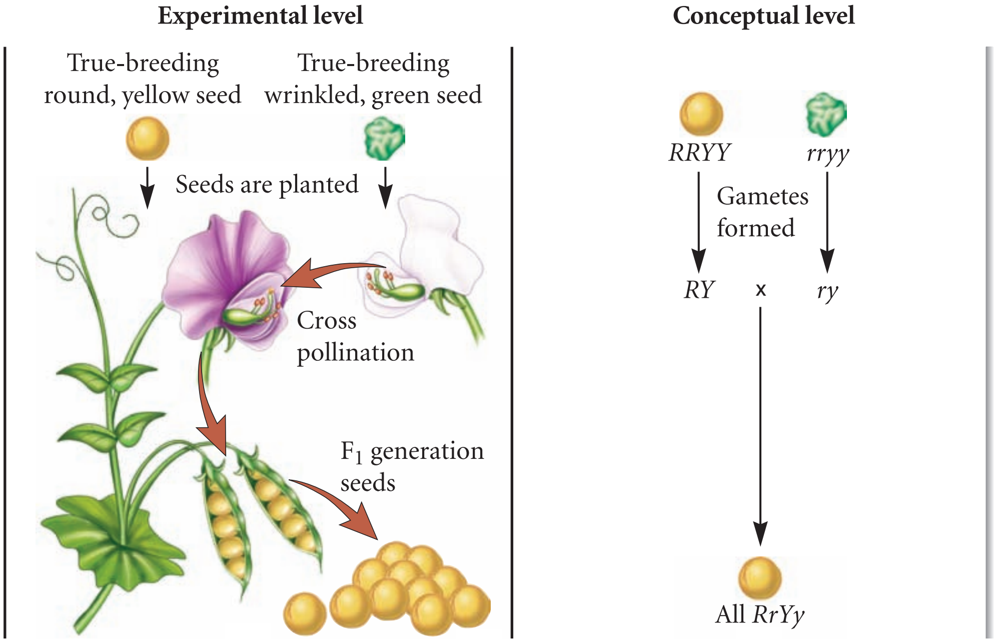
The inheritance pattern for two different characters follows one or more quantitative natural laws. Two possible hypotheses are described in Figure 3.7.

▶ TESTING THE HYPOTHESES — FIGURE 3.8 Mendel's analysis of two-factor crosses.

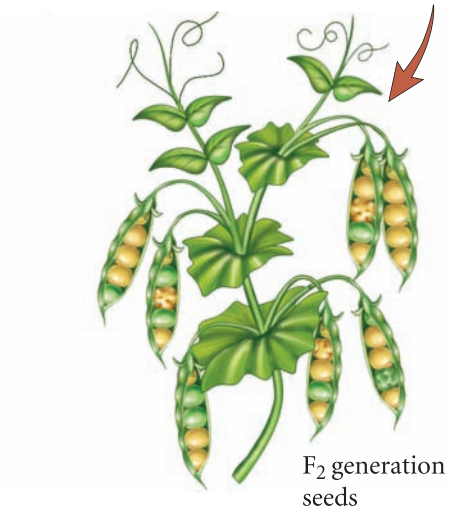
Starting material: In this experiment, Mendel began with two types of true-breeding pea strains that were different with regard to two characters. One strain produced round, yellow seeds (*RRYY*); the other strain produced wrinkled, green seeds (*rryy*).

1. Cross the two true-breeding plants to each other. This produces F_1 generation seeds.

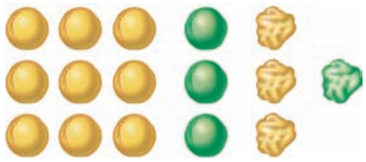
2. Collect many seeds and record their phenotype.

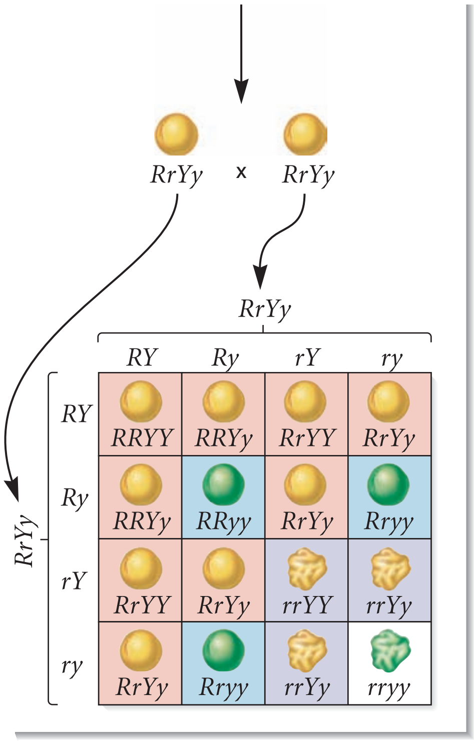

Experimental level
True-breeding round, yellow seed
True-breeding wrinkled, green seed
Seeds are planted
Cross pollination
F1 generation seeds
Conceptual level
RRYY
rryy
Gametes formed
RY
x
ry
All RrYy

3. F_1 seeds are planted and grown, and the F_1 plants are allowed to self-fertilize. This produces seeds that are part of the F_2 generation.



4. Analyze the characteristics found in the F_2 generation seeds.





	RY	Ry	rY	ry
RY	RRYY	RRYy	RrYY	RrYy
Ry	RRYy	RRyy	RrYy	Rryy
rY	RrYY	RrYy	rrYY	rrYy
ry	RrYy	Rryy	rrYy	rryy

▶ THE DATA

P cross	*F_1 generation*	*F_2 generation*
Round, yellow × wrinkled, green seeds	All round, yellow	315 round, yellow seeds 108 round, green seeds 101 wrinkled, yellow seeds 32 wrinkled, green seeds

▶ INTERPRETING THE DATA

In addition to parental-type seeds, the F_2 generation had seeds that were round and green, and seeds that were wrinkled and yellow. These two categories of F_2 seeds are called **nonparentals** because these combinations of traits were not found in the true-breeding plants of the parental generation. The occurrence of nonparental variants contradicts the linkage model. According to the linkage model, the *R* and *Y* alleles should be linked together and so should the *r* and *y* alleles. If this were the case, the F_1 plants could produce gametes that are only *RY* or *ry*. These would combine to produce *RRYY* (round, yellow), *RrYy* (round, yellow), or *rryy* (wrinkled, green) in a 1:2:1 ratio. Nonparental seeds could not be produced. However, Mendel did not obtain this result. Instead, he observed a phenotypic ratio of 9:3:3:1 in the F_2 generation.

Mendel's results from many two-factor experiments rejected the hypothesis of linked assortment and, instead, supported the hypothesis that different characters assort themselves independently. Using the modern notion of genes, **Mendel's law of independent assortment** states:

> *Two different genes randomly assort their alleles during the formation of haploid cells.*

In other words, the allele for one gene is found within a resulting gamete independently of whether the allele for a different gene is found in the same gamete. Using the example given in Figure 3.8, the round and wrinkled alleles are assorted into haploid gametes independently of the yellow and green alleles. Therefore, a heterozygous *RrYy* parent can produce four different gametes—*RY*, *Ry*, *rY*, and *ry*—in equal proportions.

In an F_1 self-fertilization experiment, any two gametes can combine randomly during fertilization. This allows for 4^2, or 16, possible offspring, although some offspring are genetically identical to each other. As shown in **Figure 3.9**, these 16 possible combinations result in seeds with the following phenotypes: 9 round, yellow; 3 round, green; 3 wrinkled, yellow; and 1 wrinkled, green. This 9:3:3:1 ratio is the expected outcome when a dihybrid is allowed to self-fertilize. Mendel was clever enough to realize that the data for his dihybrid experiments were close to a 9:3:3:1 ratio. In Figure 3.8, for example, his F_1 generation produced F_2 seeds with the following characteristics: 315 round, yellow seeds; 108 round, green seeds; 101 wrinkled, yellow seeds; and 32 wrinkled, green seeds. If we divide each of these numbers by 32 (the number of plants with wrinkled, green seeds), the phenotypic ratio of the F_2 generation is 9.8 : 3.2 : 3.4 : 1.0. Within experimental error, Mendel's data approximated the predicted 9:3:3:1 ratio for the F_2 generation.

The law of independent assortment held true for two-factor crosses involving the characters that Mendel studied in pea plants. However, in other cases, the inheritance pattern of two different genes is consistent with the linkage model described earlier in Figure 3.7a. In Chapter 7, we will examine the inheritance of genes that are linked to each other because they are physically within the same chromosome. As we will see, linked genes do not assort independently.

An important consequence of the law of independent assortment is that a single individual can produce a vast array of genetically different gametes. As discussed in Chapter 2, diploid species have pairs of homologous chromosomes, which may differ with respect to the alleles they carry. When an offspring receives a combination of alleles that differs from those in the parental generation, this phenomenon is termed **genetic recombination.** One mechanism that accounts for genetic recombination is independent assortment. A second mechanism, discussed in Chapters 2 and 7, is crossing over, which can reassort alleles that happen to be linked along the same chromosome.

The phenomenon of independent assortment is rooted in the random pattern by which the homologs assort themselves during the process of meiosis, a topic addressed later in Section 3.4. A species containing a large number of homologous chromosomes has the potential for creating an enormous amount of genetic diversity. For example, human cells contain 23 pairs of chromosomes. These pairs can randomly assort into gametes during meiosis. The number of different gametes an individual can make equals 2^n, where n is the number of pairs of chromosomes. Therefore, humans can make 2^{23}, or over 8 million, possible gametes, due to independent assortment. The capacity to make so many genetically different gametes enables a species to produce individuals with many different combinations of traits. This allows environmental factors to select for those combinations of traits that favor reproductive success.

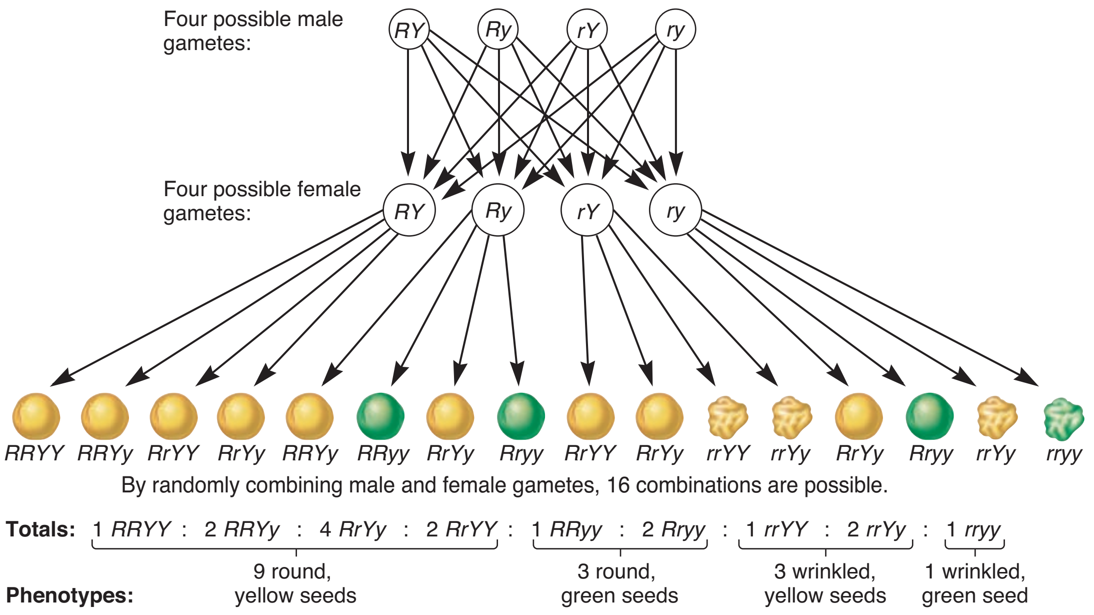
A Punnett Square Can Also Be Used to Solve Independent-Assortment Problems

As already depicted in Figure 3.8, we can make a Punnett square to predict the outcome of crosses involving two or more genes that assort independently. Let's see how such a Punnett square is made by considering a cross between two plants that are heterozygous for height and seed color (**Figure 3.10**). This cross is *TtYy* × *TtYy*. When we construct a Punnett square for this cross, we must keep in mind that each gamete has a single allele for each of two genes. In this example, the four possible gametes from each parent are

TY, *Ty*, *tY*, and *ty*

In this two-factor experiment, we need to make a Punnett square containing 16 boxes. The phenotypes of the resulting offspring are predicted to occur in a ratio of 9:3:3:1.

In crosses involving three or more genes, the construction of a single large Punnett square to predict the outcome of crosses becomes very unwieldy. For example, in a three-factor cross between two pea plants that are *Tt Rr Yy*, each parent can make 2^3, or 8, possible gametes. Therefore, the Punnett square must contain $8 \times 8 = 64$ boxes. As a more reasonable alternative, we can consider each gene separately and then algebraically combine them by multiplying together the expected outcomes for each gene. Two such methods, termed the **multiplication method** and the **forked-line method,** are shown in Solved Problem S3 at the end of this chapter.

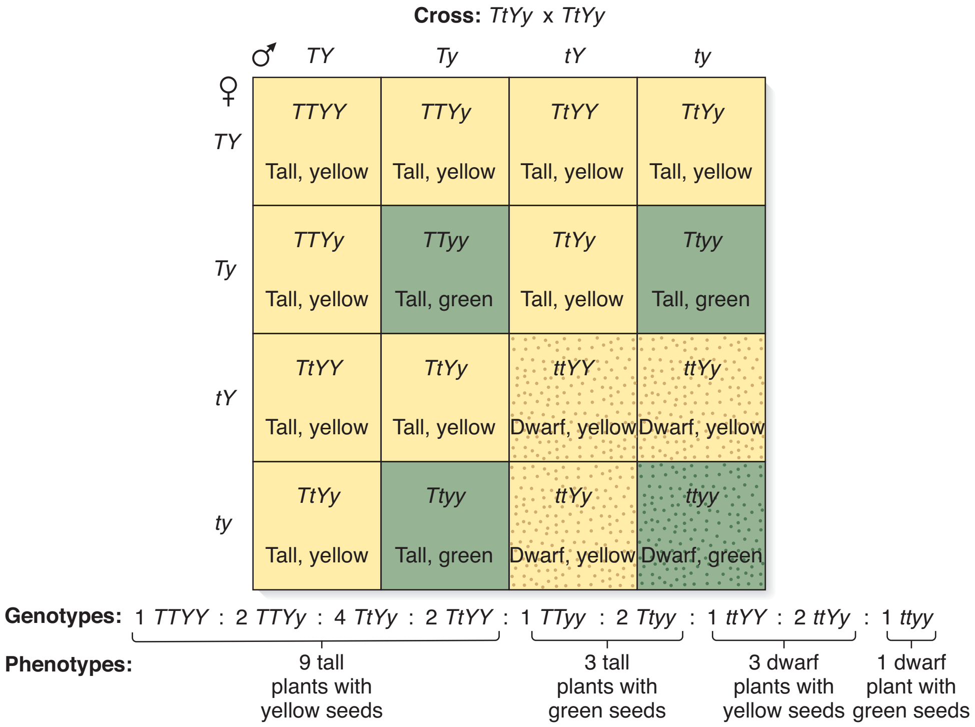

Four possible male gametes:
RY
Ry
rY
ry
Four possible female gametes:
RY
Ry
rY
ry
RRYY
RRYy
RrYY
RrYy
RRYy
RRyy
RrYy
Rryy
RrYY
RrYy
rrYY
rrYy
RrYy
Rryy
rrYy
rryy
By randomly combining male and female gametes, 16 combinations are possible.
Totals:
1 RRYY : 2 RRYy : 4 RrYy : 2 RrYY : 1 RRyy : 2 Rryy : 1 rrYY : 2 rrYy : 1 rryy
Phenotypes:
9 round, yellow seeds
3 round, green seeds
3 wrinkled, yellow seeds
1 wrinkled, green seed

ONLINE ANIMATION

FIGURE 3.9 **Mendel's law of independent assortment.**

Genes → Traits The cross is between two parents that are heterozygous for seed shape and seed color (*RrYy* × *RrYy*). Four types of male gametes are possible: *RY, Ry, rY,* and *ry*. Likewise, four types of female gametes are possible: *RY, Ry, rY,* and *ry*. These four types of gametes are the result of the independent assortment of the seed shape and seed color alleles relative to each other. During fertilization, any one of the four types of male gametes can combine with any one of the four types of female gametes. This results in 16 types of offspring, each one containing two copies of the seed shape gene and two copies of the seed color gene.

Concept check: *Why does independent assortment promote genetic variation?*



Cross: *TtYy* x *TtYy*

♀ \ ♂	*TY*	*Ty*	*tY*	*ty*
TY	*TTYY* Tall, yellow	*TTYy* Tall, yellow	*TtYY* Tall, yellow	*TtYy* Tall, yellow
Ty	*TTYy* Tall, yellow	*TTyy* Tall, green	*TtYy* Tall, yellow	*Ttyy* Tall, green
tY	*TtYY* Tall, yellow	*TtYy* Tall, yellow	*ttYY* Dwarf, yellow	*ttYy* Dwarf, yellow
ty	*TtYy* Tall, yellow	*Ttyy* Tall, green	*ttYy* Dwarf, yellow	*ttyy* Dwarf, green

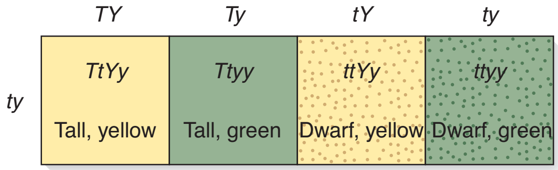
Genotypes: 1 *TTYY* : 2 *TTYy* : 4 *TtYy* : 2 *TtYY* : 1 *TTyy* : 2 *Ttyy* : 1 *ttYY* : 2 *ttYy* : 1 *ttyy*

Phenotypes: 9 tall plants with yellow seeds; 3 tall plants with green seeds; 3 dwarf plants with yellow seeds; 1 dwarf plant with green seeds

FIGURE 3.10 **A Punnett square for a two-factor cross.** The Punnett square shown here involves a cross between two pea plants that are heterozygous for height and seed color. The cross is *TtYy* × *TtYy*.

Concept check: If a parent is Ttyy, how many different types of gametes can it make?

Independent assortment is also revealed by a **two-factor testcross.** In this type of experiment, dihybrid individuals are mated to individuals that are doubly homozygous recessive for the two characters. For example, individuals with a *TtYy* genotype can be crossed to *ttyy* plants. As shown below, independent assortment predicts a 1:1:1:1 ratio among the resulting offspring.



	TY	*Ty*	*tY*	*ty*
ty	*TtYy* Tall, yellow	*Ttyy* Tall, green	*ttYy* Dwarf, yellow	*ttyy* Dwarf, green

Modern Geneticists Are Often Interested in the Relationship Between the Molecular Expression of Genes and the Outcome of Traits

Mendel's work with pea plants was critically important because his laws of inheritance pertain to most eukaryotic organisms, such as fruit flies, corn, mice, and humans, that transmit their genes through sexual reproduction. During the past several decades, many researchers have focused their attention on the relationship between the phenotypic appearance of traits and the molecular expression of genes. This theme will recur throughout the textbook (and we will draw attention to it by designating certain figure legends with a "Genes → Traits" label). As mentioned in Chapter 1, most genes encode proteins that function within living cells. The specific function of individual proteins affects the outcome of an individual's traits. A genetic approach can help us understand the relationship between a protein's function and its effect on phenotype. Most commonly, a geneticist tries to identify an individual that has a defective copy of a gene to see how the defect influences the phenotype of the organism. These defective genes are called **loss-of-function alleles,** and they provide geneticists with a great amount of information. Unknowingly, Gregor Mendel had studied seven loss-of-function alleles among his strains of plants. The recessive characteristics in his pea plants were due to genes that had been rendered defective by a mutation. Such alleles are often inherited in a recessive manner, though this is not always the case.

How are loss-of-function alleles informative? In many cases, such alleles provide critical clues about the purpose of the protein's function within the organism. For example, we expect the gene affecting flower color (purple versus white) to encode a protein that is necessary for pigment production. This protein may function as an enzyme that is necessary for the synthesis of purple pigment. Furthermore, a reasonable guess is that the white allele is a loss-of-function allele that is unable to express this protein and therefore cannot make the purple pigment. To confirm this idea, a biochemist could analyze the petals from purple and white flowers and try to identify the protein that is defective or missing in the white petals but functionally active in the purple ones. The identification and characterization of this protein provides a molecular explanation for this phenotypic characteristic.

3.3 REVIEWING THE KEY CONCEPTS

- By conducting two-factor crosses, Mendel proposed the law of independent assortment, which states that two different genes randomly assort their alleles during the formation of haploid cells (see Figures 3.8, 3.9).
- A Punnett square can be used to predict the outcome of two-factor crosses (see Figure 3.10).
- At the molecular level, the recessive variants in Mendel's crosses are due to loss-of-function mutations that prevent the expression of a functionally active protein.

3.3 COMPREHENSION QUESTIONS

1. A pea plant has the genotype *Yyrr*. How many different types of gametes can it make and in what proportions?
   a. 1 *Yy* : 1 *rr*
   b. 1 *Yr* : 1 *yr*
   c. 3 *Yr* : 1 *yr*
   d. 1 *YR* : 1 *Yr* : 1 *yR* : 1 *yr*
2. A cross is made between a plant that is *YyRr* to a plant that is *Yyrr*. What is the predicted outcome of the seed phenotypes?
   a. 9 round, yellow : 3 round, green : 3 wrinkled, yellow : 1 wrinkled green
   b. 3 round, yellow : 3 round, green : 1 wrinkled, yellow : 1 wrinkled green
   c. 3 round, yellow : 1 round, green : 3 wrinkled, yellow : 1 wrinkled green
   d. 1 round, yellow : 1 round, green : 1 wrinkled, yellow : 1 wrinkled green
3. In a population of wild squirrels, most of them have gray fur but an occasional squirrel is completely white. If we let *P* and *p* represent dominant and recessive alleles, respectively, of a gene that encodes an enzyme necessary for pigment formation, which of the following statements do you think is most likely to be correct?
   a. The white squirrels are *pp,* and the *p* allele is a loss-of-function allele.
   b. The gray squirrels are *pp,* and the *p* allele is a loss-of-function allele.
   c. The white squirrels are *PP,* and the *P* allele is a loss-of-function allele.
   d. The gray squirrels are *PP,* and the *P* allele is a loss-of-function allele.

3.4 CHROMOSOME THEORY OF INHERITANCE

Thus far, we have considered how Mendel's crosses in pea plants led to two important laws of inheritance. In Chapter 2, we examined how chromosomes are transmitted during cell division and gamete formation. In this section, we will explore how chromosomal transmission is related to the patterns of inheritance observed by Mendel. This relationship, known as the chromosome theory of inheritance, was a major breakthrough in our understanding of genetics because it established the framework for understanding how chromosomes carry and transmit the genetic determinants that govern the outcome of traits.

TABLE **3.1**

Chronology for the Development and Proof of the Chromosome Theory of Inheritance

1866	Gregor Mendel: Analyzed the transmission of traits from parents to offspring and showed that it follows a pattern of segregation and independent assortment.
1876–77	Oscar Hertwig and Hermann Fol: Observed that the nucleus of the sperm enters the egg during animal cell fertilization.
1877	Eduard Strasburger: Observed that the sperm nucleus of plants (and no detectable cytoplasm) enters the egg during plant fertilization.
1878	Walther Flemming: Described mitosis in careful detail.
1883	Carl Nägeli and August Weismann: Proposed the existence of a genetic material, which Nägeli called idioplasm and Weismann called germ plasm.
1883	Wilhelm Roux: Proposed that the most important event of mitosis is the equal partitioning of "nuclear qualities" to the daughter cells.
1883	Edouard van Beneden: showed that gametes contain half the number of chromosomes and that fertilization restores the normal diploid number.
1884–1885	Hertwig, Strasburger and August Weismann: Proposed that chromosomes are carriers of the genetic material.
1889	Theodor Boveri: Showed that enucleated sea urchin eggs that are fertilized by sperm from a different species develop into larva that have characteristics that coincide with the sperm's species.
1900	Hugo de Vries, Carl Correns, and Erich von Tschermak: Rediscovered Mendel's work, while analyzing inheritance patterns in other plants.
1901	Thomas Montgomery: Determined that maternal and paternal chromosomes pair with each other during meiosis.
1901	C. E. McClung: Discovered that sex determination in insects is related to differences in chromosome composition.
1902	Theodor Boveri: Showed that when sea urchin eggs were fertilized by two sperm, the abnormal development of the embryo was related to an abnormal number of chromosomes.
1903	Walter Sutton: Showed that even though the chromosomes seem to disappear during interphase, they do not actually disintegrate. Instead, he argued that chromosomes must retain their continuity and individuality from one cell division to the next.
1902–1903	Theodor Boveri and Walter Sutton: Independently proposed tenets of the chromosome theory of inheritance. Some historians primarily credit this theory to Sutton.
1910	Thomas Hunt Morgan: Showed that a genetic trait (i.e., white-eyed phenotype in *Drosophila*) was linked to a particular chromosome.
1913	E. Eleanor Carothers: Demonstrated that homologous pairs of chromosomes show independent assortment.
1916	Calvin Bridges: Studied chromosomal abnormalities as a way to confirm the chromosome theory of inheritance.

For a description of these experiments, the student is encouraged to read Voeller, B. R. (1968), *The Chromosome Theory of Inheritance. Classic Papers in Development and Heredity.* New York: Appleton-Century-Crofts.

This theory dramatically unfolded as a result of three lines of scientific inquiry (**Table 3.1**). One avenue concerned Mendel's breeding studies, in which he analyzed the transmission of traits from parent to offspring. A second line of inquiry involved the material basis for heredity. A Swiss botanist, Carl Nägeli, and a German biologist, August Weismann, championed the idea that a substance found in living cells is responsible for the transmission of traits from parents to offspring. Nägeli also suggested that both parents contribute equal amounts of this substance to their offspring. Several scientists, including Oscar Hertwig, Eduard Strasburger, and Walther Flemming, conducted studies suggesting that the chromosomes are the carriers of the genetic material. We now know the DNA within the chromosomes is the genetic material. Finally, the third line of evidence involved the microscopic examination of the processes of fertilization, mitosis, and meiosis. Researchers became increasingly aware that the characteristics of organisms are rooted in the continuity of cells during the life of an organism and from one generation to the next. When the work of Mendel was rediscovered, several scientists noted striking parallels between the segregation and assortment of traits noted by Mendel and the behavior of chromosomes during meiosis. Among them were Theodor Boveri, a German biologist, and Walter Sutton at Columbia University. They independently proposed the chromosome theory of inheritance, which was a milestone in our understanding of genetics.

According to the **chromosome theory of inheritance,** the inheritance patterns of traits can be explained by the transmission patterns of chromosomes during meiosis and fertilization. This theory is based on a few fundamental principles.

1. Chromosomes contain the genetic material that is transmitted from parent to offspring and from cell to cell.
2. Chromosomes are replicated and passed along, generation after generation, from parent to offspring. They are also passed from cell to cell during the development of a multicellular organism. Each type of chromosome retains its individuality during cell division and gamete formation.
3. The nuclei of most eukaryotic cells contain chromosomes that are found in homologous pairs—they are diploid. One member of each pair is inherited from the mother, the other from the father. At meiosis, one of the two members of each pair segregates into one daughter nucleus, and the homolog segregates into the other daughter nucleus. Gametes contain one set of chromosomes—they are haploid.
4. During the formation of haploid cells, different types of (nonhomologous) chromosomes segregate independently of each other.
5. Each parent contributes one set of chromosomes to its offspring. The maternal and paternal sets of homologous chromosomes are functionally equivalent; each set carries a full complement of genes.

The chromosome theory of inheritance allows us to see the relationship between Mendel's laws and chromosomal transmission. As shown in Figure 3.11, Mendel's law of segregation can be explained by the homologous pairing and segregation of chromosomes during meiosis. This figure depicts the behavior of a pair of homologous chromosomes that carry a gene for seed color. One of the chromosomes carries a dominant allele that confers yellow seed color, whereas the homologous chromosome carries a recessive allele that confers green color. A heterozygous individual passes along only one of these alleles to each offspring. In other words, a gamete may contain the yellow allele or the green allele, but not both. Because homologous chromosomes segregate from each other, a gamete contains only one copy of each type of chromosome.

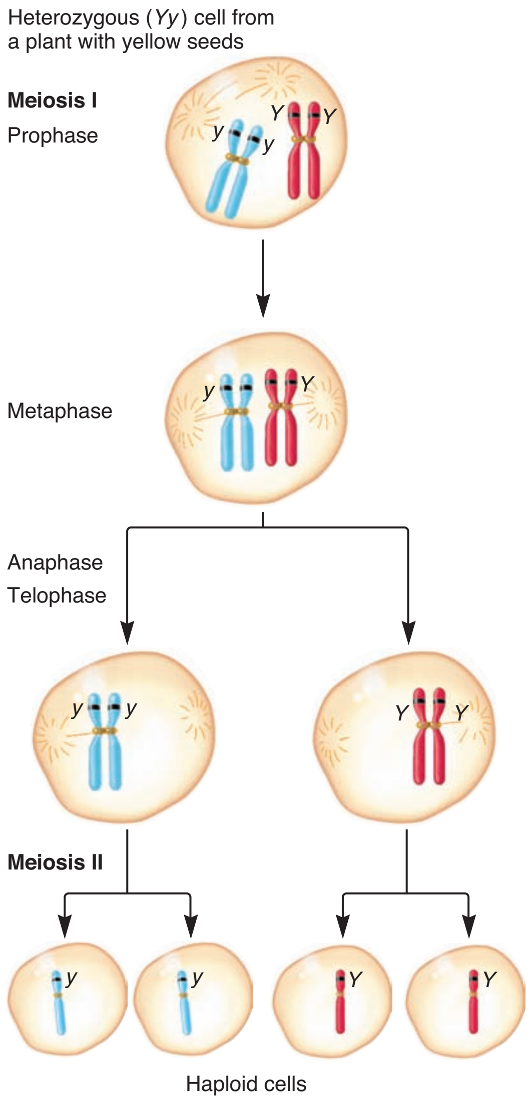


FIGURE 3.11 Mendel's law of segregation can be explained by the segregation of homologs during meiosis. The two copies of a gene are contained on homologous chromosomes. In this example using pea seed color, the two alleles are *Y* (yellow) and *y* (green). During meiosis, the homologous chromosomes segregate from each other, leading to segregation of the two alleles into separate gametes.

Genes→Traits The gene for seed color exists in two alleles, *Y* (yellow) and *y* (green). During meiosis, the homologous chromosomes that carry these alleles segregate from each other. The resulting cells receive the *Y* or *y* allele, but not both. When two gametes unite during fertilization, the alleles they carry determine the traits of the resulting offspring.

Concept check: At which stage are homologous chromosomes separating from each other?

How is the law of independent assortment explained by the behavior of chromosomes? Figure 3.12 considers the segregation of two types of chromosomes, each carrying a different gene.

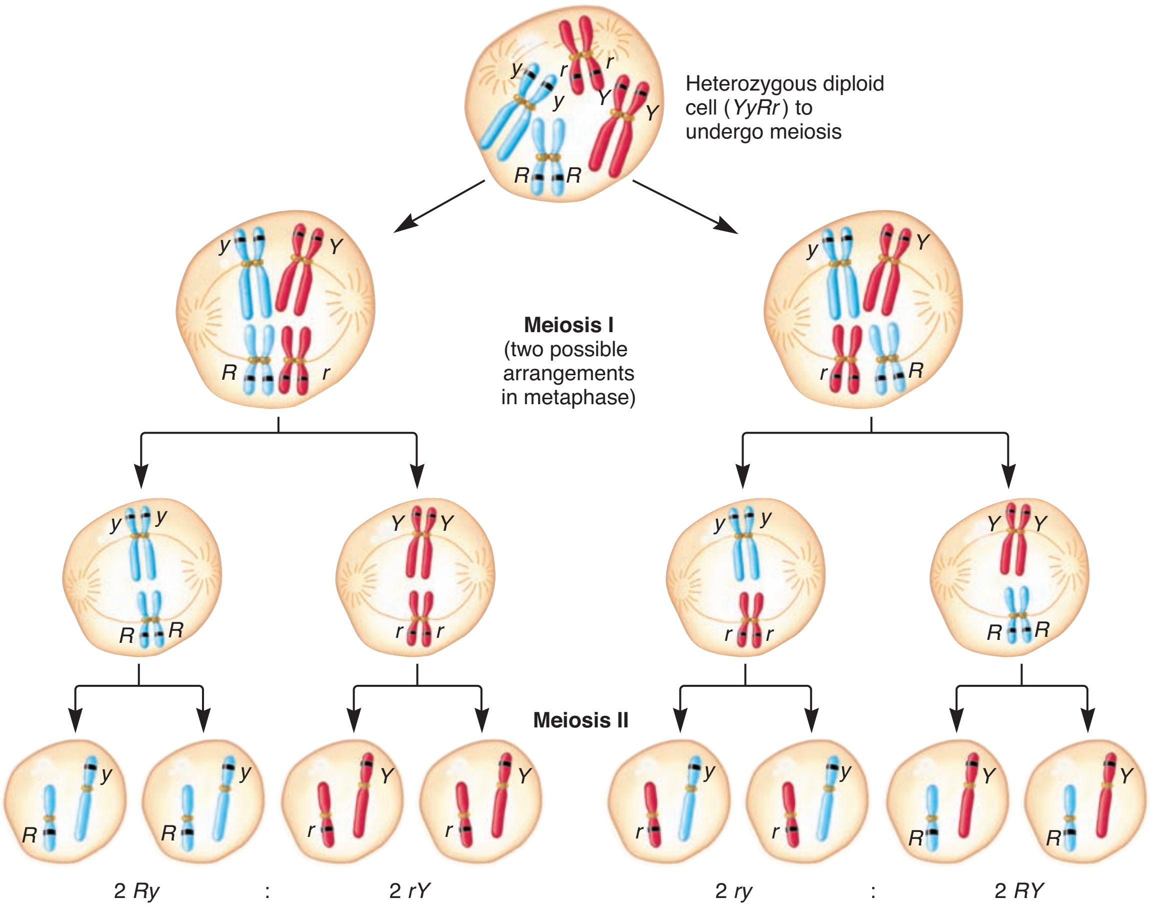

Heterozygous diploid cell (*YyRr*) to undergo meiosis
Meiosis I (two possible arrangements in metaphase)
Meiosis II
2 *Ry* : 2 *rY*
2 *ry* : 2 *RY*

FIGURE 3.12 Mendel's law of independent assortment can be explained by the random alignment of bivalents during metaphase of meiosis I. This figure shows the assortment of two genes located on two different chromosomes, using pea seed color and shape as an example (*YyRr*). During metaphase of meiosis I, different possible arrangements of the homologs within bivalents can lead to different combinations of the alleles in the resulting gametes. For example, on the left, the dominant *R* allele has sorted with the recessive *y* allele; on the right, the dominant *R* allele has sorted with the dominant *Y* allele.

Genes→Traits Most species have several different chromosomes that carry many different genes. In this example, the gene for seed color exists in two alleles, *Y* (yellow) and *y* (green), and the gene for seed shape is found as *R* (round) and *r* (wrinkled) alleles. The two genes are found on different (nonhomologous) chromosomes. During meiosis, the homologous chromosomes that carry these alleles segregate from each other. In addition, the chromosomes carrying the *Y* or *y* alleles independently assort from the chromosomes carrying the *R* or *r* alleles. As shown here, this provides a reassortment of alleles, potentially creating combinations of alleles that are different from the parental combinations. When two gametes unite during fertilization, the alleles they carry affect the traits of the resulting offspring.

Concept check: Let's suppose a pea plant is heterozygous for three genes, Tt Yy Rr, and each gene is on a different chromosome. How many different ways could the three homologous chromosomes line up during metaphase of meiosis I?

One pair of chromosomes carries the gene for seed color: the yellow (*Y*) allele is on one chromosome, and the green (*y*) allele is on the homolog. The other pair of (smaller) chromosomes carries the gene for seed shape: one copy has the round (*R*) allele, and the homolog carries the wrinkled (*r*) allele. At metaphase of meiosis I, the different types of chromosomes have randomly aligned along the metaphase plate. As shown in Figure 3.12, this can occur in more than one way. On the left, the *R* allele has sorted with the *y* allele, whereas the *r* allele has sorted with the *Y* allele. On the right, the opposite situation has occurred. Therefore, the random alignment of chromatid pairs during meiosis I can lead to an independent assortment of genes that are found on nonhomologous chromosomes. As we will see in Chapter 7, this law is violated if two different genes are located close to each other on the same chromosome.

3.4 REVIEWING THE KEY CONCEPTS

- The chromosome theory of inheritance describes how the transmission of chromosomes can explain Mendel's laws (see Table 3.1).
- Mendel's law of segregation is explained by the separation of homologs during meiosis (see Figure 3.11).
- Mendel's law of independent assortment is explained by the random alignment of different chromosomes during metaphase of meiosis I (see Figure 3.12).

3.4 COMPREHENSION QUESTIONS

1. Which of the following is *not* one of the tenets of the chromosome theory of inheritance?
   a. Chromosomes contain the genetic material that is transmitted from parent to offspring and from cell to cell.

b. Chromosomes are replicated and passed along, generation after generation, from parent to offspring.
c. Chromosome replication occurs during the S phase of the cell cycle.
d. Each parent contributes one set of chromosomes to its offspring.

2. A pea plant has the genotype *TtRr*. The independent assortment of these two genes occurs at __________ because chromosomes with the __________ alleles line up independently of the ___________ alleles.
a. metaphase of meiosis I, *T* and *t*, *R* and *r*
b. metaphase of meiosis I, *T* and *R*, *t* and *r*
c. metaphase of meiosis II, *T* and *t*, *R* and *r*
d. metaphase of meiosis II, *T* and *R*, *t* and *r*

3.5 STUDYING INHERITANCE PATTERNS IN HUMANS

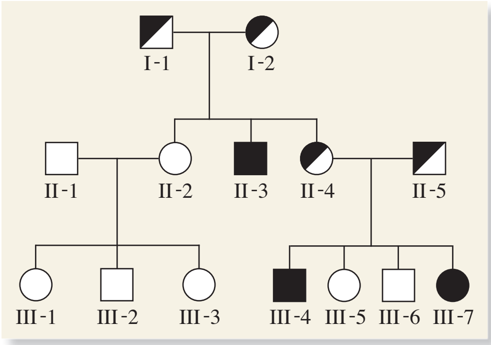
Before we end our discussion of simple Mendelian traits, let's address the question of how we can analyze inheritance patterns among humans. In his experiments, Mendel selectively made crosses and then analyzed a large number of offspring. When studying human traits, however, researchers cannot control parental crosses. Instead, they must rely on the information contained within family trees, charts representing family relationships. This type of approach, known as a **pedigree analysis,** is aimed at determining the type of inheritance pattern that a gene follows. Although this method may be less definitive than the results described in Mendel's experiments, a pedigree analysis can often provide important clues concerning the pattern of inheritance of traits within human families. An expanded discussion of human pedigrees is provided in Chapter 23, which concerns the inheritance patterns of many different human diseases.

In order to discuss the applications of pedigree analyses, we need to understand the organization and symbols of a pedigree (**Figure 3.13**). The oldest generation is at the top of the pedigree, and the most recent generation is at the bottom. Vertical lines connect each succeeding generation. A man (square) and woman (circle) who produce one or more offspring are directly connected by a horizontal line. A vertical line connects parents with their offspring. If parents produce two or more offspring, the group of siblings (brothers and sisters) is denoted by two or more individuals projecting from the same horizontal line.

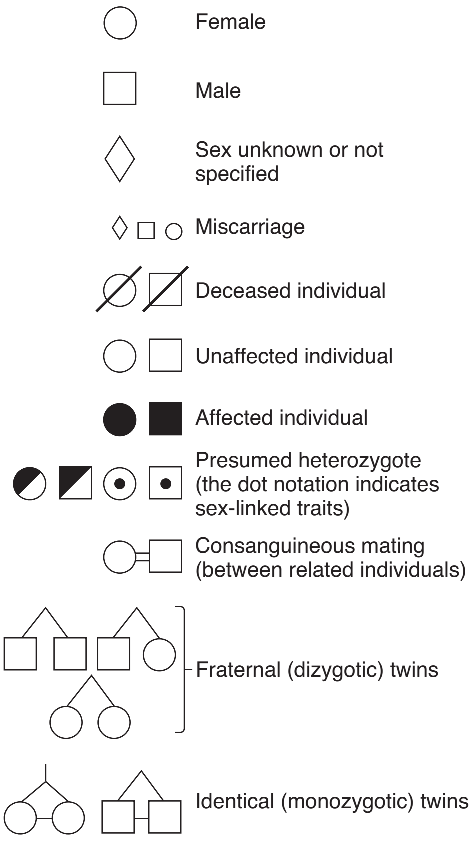
When a pedigree involves the transmission of a human trait or disease, affected individuals are depicted by filled symbols (in this case, black) that distinguish them from unaffected individuals. Each generation is given a roman numeral designation, and individuals within the same generation are numbered from left to right. A few examples of the genetic relationships in Figure 3.13a are described here:

> Individuals I-1 and I-2 are the grandparents of III-1, III-2, III-3, III-4, III-5, III-6, and III-7
>
> Individuals III-1, III-2, and III-3 are brother and sisters
>
> Individuals II-3, III-4, and III-7 are affected by a genetic disease

The symbols shown in Figure 3.13 depict certain individuals, such as I-1, I-2, II-4, and II-5, as presumed heterozygotes because they are unaffected with a disease but produce homozygous offspring that are affected with a recessive genetic disease. However, in many pedigrees, such as those found in the problem



(a) Human pedigree showing cystic fibrosis


Female
Male
Sex unknown or not specified
Miscarriage
Deceased individual
Unaffected individual
Affected individual
Presumed heterozygote (the dot notation indicates sex-linked traits)
Consanguineous mating (between related individuals)
Fraternal (dizygotic) twins
Identical (monozygotic) twins

(b) Symbols used in a human pedigree

FIGURE 3.13 **Pedigree analysis. (a)** A family pedigree in which some of the members are affected with cystic fibrosis. Individuals I-1, I-2, II-4, and II-5 are depicted as presumed heterozygotes because they produce affected offspring. **(b)** The symbols used in a pedigree analysis. *Note:* In most pedigrees shown in this textbook, such as those found in the problem sets, the heterozygotes are not shown as half-filled symbols. Most pedigrees throughout the book show individuals' phenotypes—open symbols are unaffected individuals and filled (closed) symbols are affected individuals.

Concept check: *What are the two different meanings of horizontal lines in a pedigree?*

sets at the end of the chapter, the inheritance pattern may not be known, so the symbols reflect only phenotypes. In most pedigrees, affected individuals are shown with closed symbols, and unaffected individuals, including those that might be heterozygous for a recessive disease, are depicted with open symbols.

Pedigree analysis is commonly used to determine the inheritance pattern of human genetic diseases. Human geneticists are routinely interested in knowing whether a genetic disease is inherited as a recessive or dominant trait. One way to discern the dominant/recessive relationship between two alleles is by a pedigree analysis. Genes that play a role in disease may exist as a normal allele or a mutant allele that causes disease symptoms. If the disease follows a simple Mendelian pattern of inheritance and is caused by a recessive allele, an individual must inherit two copies of the mutant allele to exhibit the disease. Therefore, a recessive pattern of inheritance makes two important predictions. First, two heterozygous normal individuals will, on average, have 1/4 of their offspring affected. Second, all offspring of two affected individuals will be affected. Alternatively, a dominant trait predicts that affected individuals will have inherited the gene from at least one affected parent unless a new mutation has occurred during gamete formation.

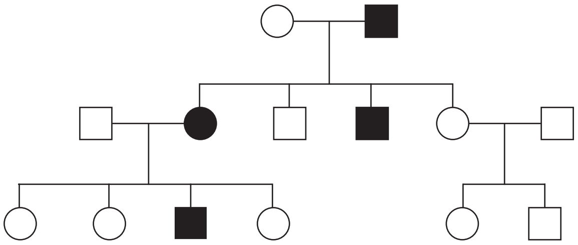
The pedigree in Figure 3.13a concerns a human genetic disease known as cystic fibrosis (CF). Among Caucasians, approximately 3% of the population are heterozygous carriers of this recessive allele. In homozygotes, the disease symptoms include abnormalities of the pancreas, intestine, sweat glands, and lungs. These abnormalities are caused by an imbalance of ions across the plasma membrane. In the lungs, this leads to a buildup of thick, sticky mucus. Respiratory problems may lead to early death, although modern treatments have greatly increased the life span of CF patients. In the late 1980s, the gene for CF was identified. The *CFTR* gene encodes a protein called the cystic fibrosis transmembrane conductance regulator (CFTR). This protein regulates the ion balance across the cell membrane in tissues of the pancreas, intestine, sweat glands, and lungs. The mutant allele causing CF alters the encoded CFTR protein. The altered CFTR protein is not correctly inserted into the plasma membrane, resulting in a decreased function that causes the ionic imbalance. As seen in the pedigree, the pattern of affected and unaffected individuals is consistent with a recessive mode of inheritance. Two unaffected individuals can produce an affected offspring. Although not shown in this pedigree, a recessive mode of inheritance is also characterized by the observation that two affected individuals will produce 100% affected offspring. However, for human genetic diseases that limit survival or fertility (or both), there may never be cases where two affected individuals produce offspring.

3.5 REVIEWING THE KEY CONCEPTS

- Human inheritance patterns are determined by analyzing family trees known as pedigrees (see Figure 3.13).

3.5 COMPREHENSION QUESTIONS

1. Which of the following would *not* be observed in a pedigree if a genetic disorder was inherited in a recessive manner?
   a. Two unaffected parents have an affected offspring.
   b. Two affected parents have an unaffected offspring.
   c. One affected and one unaffected parent have an unaffected offspring.
   d. All of the above are possible for a recessive disorder.
2. In the pedigree shown here, which pattern(s) of inheritance is/are possible? Affected individuals are shown with a filled symbol.



   a. Recessive
   b. Dominant
   c. Both recessive and dominant
   d. Neither recessive or dominant

3.6 PROBABILITY AND STATISTICS

A powerful application of Mendel's work is that the laws of inheritance can be used to predict the outcome of genetic crosses. In agriculture, for example, plant and animal breeders are concerned with the types of offspring their crosses will produce. This information is used to produce commercially important crops and livestock. In addition, people are often interested in predicting the characteristics of the children they may have. This may be particularly important to individuals who carry alleles that cause inherited diseases. Of course, we cannot see into the future and definitively predict what will happen. Nevertheless, genetic counselors can help couples to predict the likelihood of having an affected child. This probability is one factor that may influence a couple's decision to have children.

In this section, we will see how probability calculations are used in genetic problems to predict the outcome of crosses. To compute probability, we will use three mathematical operations known as the sum rule, the product rule, and the binomial expansion equation. These methods allow us to determine the probability that a cross between two individuals will produce a particular outcome. To apply these operations, we must have knowledge regarding the genotypes of the parents and the pattern of inheritance of a given trait.

Probability calculations can also be used in hypothesis testing. In many situations, a researcher wants to discern the genotypes and patterns of inheritance for traits that are not yet understood. A traditional approach to this problem is to conduct crosses and then analyze their outcomes. The proportions of offspring may provide important clues that allow the experimenter to propose a hypothesis, based on the quantitative laws of inheritance, that explains the transmission of the trait from parent to offspring. Statistical methods, such as the chi square test, can then be used to evaluate how well the observed data from crosses fit the expected data. We will end this chapter with an example that applies the chi square test to a genetic cross.

Probability Is the Likelihood That an Event Will Occur

The chance that an event will occur in the future is called the event's **probability.** For example, if you flip a coin, the probability is 0.50, or 50%, that the head side will be showing when the coin lands. Probability depends on the number of possible outcomes. In this case, two possible outcomes (heads or tails) are equally likely. This allows us to predict a 50% chance that a coin flip will produce heads. The general formula for probability (P) is

$$\text{Probability} = \frac{\text{Number of times an event will occur}}{\text{Total number of possible events}}$$

$$P_{\text{heads}} = 1 \text{ heads} / (1 \text{ heads} + 1 \text{ tails}) = 1/2 = 50\%$$

In genetic problems, we are often interested in the probability that a particular type of offspring will be produced. Recall that when two heterozygous tall pea plants (*Tt*) are crossed, the phenotypic ratio of the offspring is 3 tall to 1 dwarf. This information can be used to calculate the probability for either type of offspring.

$$\text{Probability} = \frac{\begin{array}{c}\text{Expected number of individuals}\\ \text{with a given phenotype}\end{array}}{\text{Total number of individuals}}$$

$$P_{\text{tall}} = 3 \text{ tall} / (3 \text{ tall} + 1 \text{ dwarf}) = 3/4 = 75\%$$

$$P_{\text{dwarf}} = 1 \text{ dwarf} / (3 \text{ tall} + 1 \text{ dwarf}) = 1/4 = 25\%$$

The probability is 75% of obtaining a tall plant and 25% of obtaining a dwarf plant. When we add together the probabilities of all possible outcomes (tall and dwarf), we should get a sum of 100% (here, 75% + 25% = 100%).

A probability calculation allows us to predict the likelihood that an event will occur in the future. The accuracy of this prediction, however, depends to a great extent on the size of the sample. For example, if we toss a coin six times, our probability prediction suggests that 50% of the time we should get heads (i.e., three heads and three tails). In this small sample size, however, we would not be too surprised if we came up with four heads and two tails. Each time we toss a coin, there is a random chance that it will be heads or tails. The deviation between the observed and expected outcomes is called the **random sampling error.** In a small sample, the error between the predicted percentage of heads and the actual percentage observed may be quite large. By comparison, if we flipped a coin 1000 times, the percentage of heads would be fairly close to the predicted 50% value. In a larger sample, we expect the random sampling error to be a much smaller percentage.

The Sum Rule Can Be Used to Predict the Occurrence of Mutually Exclusive Events

Now that we are familiar with the concept of probability, let's consider how mathematical operations using probability values allow us to predict the outcome of genetic crosses. Our first genetic problem involves the use of the **sum rule,** which states:

> *The probability that one of two or more mutually exclusive events will occur is equal to the sum of the individual probabilities of the events.*

As an example, let's consider a cross between two mice that are both heterozygous for genes affecting the ears and tail. One gene can be found as an allele designated *de*, which is a recessive allele that causes droopy ears; the normal allele is *De*. An allele of a second gene causes a crinkly tail. This crinkly tail allele (*ct*) is recessive to the normal allele (*Ct*). If a cross is made between two heterozygous mice (*Dede Ctct*), the predicted ratio of offspring is 9 with normal ears and normal tails, 3 with normal ears and crinkly tails, 3 with droopy ears and normal tails, and 1 with droopy ears and a crinkly tail. These four phenotypes are mutually exclusive. For example, a mouse with droopy ears and a normal tail cannot have normal ears and a crinkly tail.

The sum rule allows us to determine the probability that we will obtain any one of two or more different types of offspring. For example, in a cross between two heterozygotes (*Dede Ctct* × *Dede Ctct*), we can ask the following question: What is the probability that an offspring will have normal ears and a normal tail or have droopy ears and a crinkly tail? In other words, if we closed our eyes and picked an offspring out of a litter from this cross, what are the chances that we would be holding a mouse that has normal ears and a normal tail or a mouse with droopy ears and a crinkly tail? In this case, the investigator wants to predict whether one of two mutually exclusive events will occur. A strategy for solving such genetic problems using the sum rule is described here.

The Cross: *Dede Ctct* × *Dede Ctct*

The Question: What is the probability that an offspring will have normal ears and a normal tail or have droopy ears and a crinkly tail?

Step 1. *Calculate the individual probabilities of each phenotype.* This can be accomplished using a Punnett square.

The probability of normal ears and a normal tail is $9/(9 + 3 + 3 + 1) = 9/16$

The probability of droopy ears and a crinkly tail is $1/(9 + 3 + 3 + 1) = 1/16$

Step 2. *Add together the individual probabilities.*

$$9/16 + 1/16 = 10/16$$

This means that 10/16 is the probability that an offspring will have either normal ears and a normal tail or droopy ears and a crinkly tail. We can convert 10/16 to 0.625, which means that 62.5% of the offspring are predicted to have normal ears and a normal tail or droopy ears and a crinkly tail.

The Product Rule Can Be Used to Predict the Probability of Independent Events

We can use probability to make predictions regarding the likelihood of two or more independent outcomes from a genetic cross. When we say that events are independent, we mean that the occurrence of one event does not affect the probability of another event. As an example, let's consider a rare, recessive human trait known as congenital analgesia. Persons with this trait can distinguish between sharp and dull, and hot and cold, but do not perceive extremes of sensation as being painful. The first case of congenital analgesia, described in 1932, was a man who made his living entertaining the public as a "human pincushion."

For a phenotypically unaffected couple, each being heterozygous for the recessive allele causing congenital analgesia, we can ask the question, What is the probability that the couple's first three offspring will have congenital analgesia? To answer this question, we can use the **product rule** which states:

> *The probability that two or more independent events will occur is equal to the product of their individual probabilities.*

A strategy for solving this type of problem is shown here.

The Cross: *Pp* × *Pp* (where *P* is the common allele and *p* is the recessive congenital analgesia allele)

The Question: What is the probability that the couple's first three offspring will have congenital analgesia?

Step 1. *Calculate the individual probability of this phenotype.* As described previously, this is accomplished using a Punnett square.

The probability of an affected offspring is 1/4 (25%).

Step 2. *Multiply the individual probabilities.* In this case, we are asking about the first three offspring, and so we multiply 1/4 three times.

$$1/4 \times 1/4 \times 1/4 = 1/64 = 0.016$$

Thus, the probability that the first three offspring will have this trait is 0.016. In other words, we predict that 1.6% of the time the first three offspring of a couple, each heterozygous for the recessive allele, will all have congenital analgesia. In this example, the phenotypes of the first, second, and third offspring are independent events. The phenotype of the first offspring does not have an effect on the phenotype of the second or third offspring.

In the problem described here, we have used the product rule to determine the probability that the first three offspring will all have the same phenotype (congenital analgesia). We can also apply the rule to predict the probability of a sequence of events that involves combinations of different offspring. For example, consider the question, What is the probability that the first offspring will be unaffected, the second offspring will have congenital analgesia, and the third offspring will be unaffected? Again, to solve this problem, begin by calculating the individual probability of each phenotype.

Unaffected = 3/4
Congenital analgesia = 1/4

The probability that these three phenotypes will occur in this specified order is

$$3/4 \times 1/4 \times 3/4 = 9/64 = 0.14, \text{ or } 14\%$$

In other words, this sequence of events is expected to occur only 14% of the time.

The product rule can also be used to predict the outcome of a cross involving two or more genes. Let's suppose an individual with the genotype *Aa Bb CC* was crossed to an individual with the genotype *Aa bb Cc*. We could ask the question, What is the probability that an offspring will have the genotype *AA bb Cc*? If the three genes independently assort, the probability of inheriting alleles for each gene is independent of the other two genes. Therefore, we can separately calculate the probability of the desired outcome for each gene.

Cross: *Aa Bb CC* × *Aa bb Cc*

Probability that an offspring will be *AA* = 1/4, or 0.25 (25%)

Probability that an offspring will be *bb* = 1/2, or 0.5 (50%)

Probability that an offspring will be *Cc* = 1/2, or 0.5 (50%)

We can use the product rule to determine the probability that an offspring will be *AA bb Cc;*

$$P = (0.25)(0.5)(0.5) = 0.0625, \text{ or } 6.25\%$$

The Binomial Expansion Equation Can Be Used to Predict the Probability of an Unordered Combination of Events

A third predictive problem in genetics is to determine the probability that a certain proportion of offspring will be produced with particular characteristics; here they can be produced in an unspecified order. For example, we can consider a group of children produced by two heterozygous brown-eyed (*Bb*) individuals. We can ask the question, What is the probability that two out of five children will have blue eyes?

In this case, we are not concerned with the order in which the offspring are born. Instead, we are only concerned with the final numbers of blue-eyed and brown-eyed offspring. One possible outcome would be the following: firstborn child with blue eyes, second child with blue eyes, and then the next three with brown eyes. Another possible outcome could be firstborn child with brown eyes, second with blue eyes, third with brown eyes, fourth with blue eyes, and fifth with brown eyes. Both of these scenarios would result in two offspring with blue eyes and three with brown eyes. In fact, several other ways to have such a family could occur.

To solve this type of question, the **binomial expansion equation** can be used. This equation represents all of the possibilities for a given set of unordered events.

$$P = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

where

P = the probability that the unordered outcome will occur
n = total number of events
x = number of events in one category (e.g., blue eyes)
p = individual probability of x
q = individual probability of the other category (e.g., brown eyes)

Note: In this case, $p + q = 1$.

The symbol ! denotes a factorial. *n*! is the product of all integers from *n* down to 1. For example, $4! = 4 \times 3 \times 2 \times 1 = 24$. An exception is 0!, which equals 1.

The use of the binomial expansion equation is described next.

The Cross: *Bb* × *Bb*

The Question: What is the probability that two out of five offspring will have blue eyes?

Step 1. *Calculate the individual probabilities of the blue-eye and brown-eye phenotypes.* If we construct a Punnett square, we find the probability of blue eyes is 1/4 and the probability of brown eyes is 3/4:

$$p = 1/4$$

$$q = 3/4$$

Step 2. *Determine the number of events in category x (in this case, blue eyes) versus the total number of events.* In this example, the number of events in category *x* is two blue-eyed children among a total number of five.

$$x = 2$$

$$n = 5$$

Step 3. *Substitute the values for p, q, x, and n in the binomial expansion equation.*

$$P = \frac{n!}{x!(n-x)!}\, p^x q^{n-x}$$

$$P = \frac{5!}{2!(5-2)!}\,(1/4)^2(3/4)^{5-2}$$

$$P = \frac{5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(3 \times 2 \times 1)}\,(1/16)(27/64)$$

$$P = 0.26 = 26\%$$

Thus, the probability is 0.26 that two out of five offspring will have blue eyes. In other words, 26% of the time we expect a *Bb* × *Bb* cross yielding five offspring to contain two blue-eyed children and three brown-eyed children.

In Solved Problem S6 at the end of this chapter, we consider an expanded version of this approach that uses a **multinomial expansion equation.** This equation is needed to solve unordered genetic problems that involve three or more phenotypic categories.

The Chi Square Test Can Be Used to Test the Validity of a Genetic Hypothesis

We now look at a different issue in genetic problems, namely **hypothesis testing.** Our goal here is to determine if the data from genetic crosses are consistent with a particular pattern of inheritance. For example, a geneticist may study the inheritance of body color and wing shape in fruit flies over the course of two generations. The following question may be asked about the F_2 generation: Do the observed numbers of offspring agree with the predicted numbers based on Mendel's laws of segregation and independent assortment? As we will see in Chapters 4 through 8, not all traits follow a simple Mendelian pattern of inheritance. Some genes do not segregate and independently assort themselves in the same way that Mendel's seven characters did in pea plants.

To distinguish between inheritance patterns that obey Mendel's laws versus those that do not, a conventional strategy is to make crosses and then quantitatively analyze the offspring. Based on the observed outcome, an experimenter may make a tentative hypothesis. For example, it may seem that the data are obeying Mendel's laws. Hypothesis testing provides an objective, statistical method to evaluate whether the observed data really agree with the hypothesis. In other words, we use statistical methods to determine whether the data that have been gathered from crosses are consistent with predictions based on quantitative laws of inheritance.

The rationale behind a statistical approach is to evaluate the **goodness of fit** between the observed data and the data that are predicted from a hypothesis. This is called a **null hypothesis** when it assumes there is no real difference between the observed and expected values. Any actual differences that occur are presumed to be due to random sampling error. If the observed and predicted data are very similar, we can conclude that the hypothesis is consistent with the observed outcome. In this case, it is reasonable to accept the null hypothesis. However, it should be emphasized that this does not prove a hypothesis is correct. Statistical methods can never prove a hypothesis is correct. They can provide insight about whether or not the observed data seem reasonably consistent with the hypothesis. Alternative hypotheses, perhaps even ones that the experimenter has failed to realize, may also be consistent with the data. In some cases, statistical methods may reveal a poor fit between hypothesis and data. In other words, a high deviation is found between the observed and expected values. If this occurs, the null hypothesis is rejected. Hopefully, the experimenter can subsequently propose an alternative hypothesis that has a better fit with the data.

One commonly used statistical method for determining goodness of fit is the **chi square test** (often written χ^2). We can use the chi square test to analyze population data in which the members of the population fall into different categories. This is the kind of data we have when we evaluate the outcome of genetic crosses, because these usually produce a population of offspring that differ with regard to phenotypes. The general formula for the chi square test is

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$

where

O = observed data in each category

E = expected data in each category based on the experimenter's hypothesis

Σ means to sum this calculation for each category. For example, if the population data fell into two categories, the chi square calculation would be

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

We can use the chi square test to determine if a genetic hypothesis is consistent with the observed outcome of a genetic cross. The strategy described next provides a step-by-step outline for applying the chi square testing method. In this problem, the experimenter wants to determine if a two-factor cross is obeying Mendel's laws. The experimental organism is *Drosophila melanogaster* (the common fruit fly), and the two characters involve wing shape and body color. Straight wing shape and curved wing shape are designated by c^+ and c, respectively; gray body color and ebony body color are designated by e^+ and e, respectively. Note: In certain species, such as *Drosophila melanogaster,* the convention is to designate the common (wild-type) allele with a plus sign. Recessive mutant alleles are designated with lowercase letters and dominant mutant alleles with capital letters.

The Cross: A true-breeding fly with straight wings and a gray body ($c^+c^+e^+e^+$) is crossed to a true-breeding fly with curved wings and an ebony body (*ccee*). The flies of the F_1 generation are then allowed to mate with each other to produce an F_2 generation.

The Outcome:

F_1 generation:	All offspring have straight wings and gray bodies
F_2 generation:	193 straight wings, gray bodies
	69 straight wings, ebony bodies
	64 curved wings, gray bodies
	26 curved wings, ebony bodies
Total:	352

Step 1. *Propose a hypothesis that allows us to calculate the expected values based on Mendel's laws.* The F_1 generation suggests that the trait of straight wings is dominant to curved wings and gray body coloration is dominant to ebony. Looking at the F_2 generation, it appears that the data are following a 9:3:3:1 ratio. If so, this is consistent with an independent assortment of the two characters.

Based on these observations, the hypothesis is:

Straight (c^+) is dominant to curved (c), and gray (e^+) is dominant to ebony (e). The two characters segregate and assort independently from generation to generation.

Step 2. *Based on the hypothesis, calculate the expected values of the four phenotypes.* We first need to calculate the individual probabilities of the four phenotypes. According to our hypothesis, there should be a 9:3:3:1 ratio in the F_2 generation. Therefore, the expected probabilities are:

$9/16$ = straight wings, gray bodies
$3/16$ = straight wings, ebony bodies
$3/16$ = curved wings, gray bodies
$1/16$ = curved wings, ebony bodies

The observed F_2 generation contained a total of 352 individuals. Our next step is to calculate the expected numbers of each type of offspring when the total equals 352. This can be accomplished by multiplying each individual probability by 352.

$9/16 \times 352 = 198$ (expected number with straight wings, gray bodies)
$3/16 \times 352 = 66$ (expected number with straight wings, ebony bodies)
$3/16 \times 352 = 66$ (expected number with curved wings, gray bodies)
$1/16 \times 352 = 22$ (expected number with curved wings, ebony bodies)

Step 3. *Apply the chi square formula, using the data for the expected values that have been calculated in step 2.* In this case, the data include four categories, and thus the sum has four terms.

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$\chi^2 = \frac{(193 - 198)^2}{198} + \frac{(69 - 66)^2}{66} + \frac{(64 - 66)^2}{66} + \frac{(26 - 22)^2}{22}$$

$$\chi^2 = 0.13 + 0.14 + 0.06 + 0.73 = 1.06$$

Step 4. *Interpret the calculated chi square value. This is done using a chi square table.*

Before interpreting the chi square value we obtained, we must understand how to use **Table 3.2.** The probabilities, called ***P* values,** listed in the chi square table allow us to determine the likelihood that the amount of variation indicated by a given chi square value is due to random chance alone, based on a particular hypothesis. For example, let's consider a value (0.00393) listed in row 1. (The meaning of the rows will be explained shortly.) Chi square values that are equal to or greater than 0.00393 are expected to occur 95% of the time when a hypothesis is correct. In other words, 95 out of 100 times we would expect that random chance alone would produce a deviation between the experimental data and hypothesized model that is equal to or greater than 0.00393. A low chi square value indicates a high probability that the observed deviation could be due to random chance alone. By comparison, chi square values that are equal to or greater than 3.841 are expected to occur less than 5% of the time due to random sampling error. If a high chi square value is obtained, an experimenter becomes suspicious that the high deviation has occurred because the null hypothesis is incorrect and that the deviation is not attributed to chance alone.

A common convention is to reject the null hypothesis if the chi square value results in a probability that is less than 0.05 (less than 5%) or if the probability is less than 0.01 (less than 1%). These are called the 5% and 1% significance levels, respectively. Which level is better to chose? The choice is somewhat subjective. If you chose a 5% level rather than a 1% level, a disadvantage is that you are more likely to reject a hypothesis that happens to be correct. Even so, choosing a 5% level rather than a 1% level has the advantage that you are less likely to accept an incorrect hypothesis.

In our problem involving flies with straight or curved wings and gray or ebony bodies, we have calculated a chi square value of 1.06. Before we can determine the probability that this deviation would have occurred as a matter of random chance, we must first determine the degrees of freedom (*df*) in this

TABLE 3.2

Chi Square Values and Probability

						Null Hypothesis Rejected	
Degrees of Freedom	$P = 0.99$	0.95	0.80	0.50	0.20	0.05	0.01
1.	0.000157	0.00393	0.0642	0.455	1.642	3.841	6.635
2.	0.020	0.103	0.446	1.386	3.219	5.991	9.210
3.	0.115	0.352	1.005	2.366	4.642	7.815	11.345
4.	0.297	0.711	1.649	3.357	5.989	9.488	13.277
5.	0.554	1.145	2.343	4.351	7.289	11.070	15.086
6.	0.872	1.635	3.070	5.348	8.558	12.592	16.812
7.	1.239	2.167	3.822	6.346	9.803	14.067	18.475
8.	1.646	2.733	4.594	7.344	11.030	15.507	20.090
9.	2.088	3.325	5.380	8.343	12.242	16.919	21.666
10.	2.558	3.940	6.179	9.342	13.442	18.307	23.209
15.	5.229	7.261	10.307	14.339	19.311	24.996	30.578
20.	8.260	10.851	14.578	19.337	25.038	31.410	37.566
25.	11.524	14.611	18.940	24.337	30.675	37.652	44.314
30.	14.953	18.493	23.364	29.336	36.250	43.773	50.892

From Fisher, R. A., and Yates, F. (1943) *Statistical Tables for Biological, Agricultural, and Medical Research.* Oliver and Boyd, London.

experiment. The **degrees of freedom** is a measure of the number of categories that are independent of each other. When phenotype categories are derived from a Punnett square, it is typically $n - 1$, where n equals the total number of categories. In the preceding problem, $n = 4$ (the categories are the phenotypes: straight wings and gray body; straight wings and ebony body; curved wings and gray body; and curved wings and ebony body); thus, the degrees of freedom equals 3.[*] We now have sufficient information to interpret our chi square value of 1.06.

With $df = 3$, the chi square value of 1.06 we obtained is slightly greater than 1.005, which gives a P value of 0.80, or 80%. What does this P value mean? If the hypothesis is correct, chi square values equal to or greater than 1.005 are expected to occur 80% of the time based on random chance alone. To reject the null hypothesis at the 5% significance level, the chi square would have to be greater than 7.815. Because it was actually far less than this value, we are inclined to accept that the hypothesis is correct.

Again, keep in mind that the chi square test does not prove a hypothesis is correct. It is a statistical method for evaluating whether the data and hypothesis have a good fit.

3.6 REVIEWING THE KEY CONCEPTS

- Probability is the number of times an event occurs divided by the total number of events.
- According to the sum rule, the probability that one of two or more mutually exclusive events will occur is equal to the sum of the individual probabilities of the events.
- According to the product rule, the probability of two or more independent events is equal to the product of their individual probabilities. This rule can be used to predict the outcome of crosses involving two or more genes.
- The binomial expansion is used to predict the probability of an unordered combination of events.
- The chi square test is used to test the validity of a hypothesis (see Table 3.1).

3.6 COMPREHENSION QUESTIONS

1. A cross is made between *AA Bb Cc Dd* and *Aa Bb cc dd* individuals. Rather than making a very large Punnett square, which statistical operation could you use to solve this problem, and what would be the probability of an offspring that is *AA bb Cc dd*?
   a. Product rule, 1/32
   b. Product rule, 1/4
   c. Sum rule, 1/32
   d. Sum rule, 1/4
2. In dogs, brown fur color (*B*) is dominant to white (*b*). A cross is made between two heterozygotes. If a litter contains six pups, what is the probability that half of them will be white?
   a. 0.066, or 6.6%
   b. 0.13, or 13%
   c. 0.25, or 25%
   d. 0.26, or 26%
3. Which of the following operations could be used for hypothesis testing?
   a. Sum rule
   b. Binomial expansion
   c. Product rule
   d. Chi square test

---

[*] If our hypothesis already assumed that the law of segregation is obeyed, the degrees of freedom would be 1 (see Chapter 7).

KEY TERMS

Page 40. pangenesis, blending hypothesis of inheritance
Page 41. crossed, hybridization, hybrids
Page 42. gamete, sperm, pollen grains, anthers, eggs, ovules, ovaries, stigma
Page 43. self-fertilization, cross-fertilization, characters, trait, variant, true-breeding line, strain
Page 44. single-factor cross, monohybrids, empirical approach, empirical laws, parental generation, P generation, F_1 generation, F_2 generation
Page 46. dominant, recessive, particulate theory of inheritance, segregate, gene, allele, Mendel's law of segregation, homozygous, genotype, phenotype, heterozygous
Page 47. Punnett square
Page 48. two-factor crosses
Page 50. nonparentals, Mendel's law of independent assortment
Page 51. genetic recombination, multiplication method, forked-line method
Page 52. two-factor testcross, loss-of-function alleles
Page 54. chromosome theory of inheritance
Page 56. pedigree analysis
Page 58. probability, random sampling error, sum rule
Page 59. product rule, binomial expansion equation
Page 60. multinomial expansion equation, hypothesis testing, goodness of fit, null hypothesis, chi square test
Page 61. *P* values
Page 62. degrees of freedom

CHAPTER SUMMARY

- Early ideas regarding inheritance included pangenesis and blending inheritance. These ideas were later refuted by the work of Mendel.

3.1 Mendel's Study of Pea Plants

- Mendel chose pea plants as his experimental organism (see Figures 3.1, 3.2).
- Using pea plants, it was easy for Mendel to carry out self-fertilization or cross-fertilization experiments (see Figure 3.3).
- Another advantage of using pea plants is they were available in several varieties in which a character existed in two distinct variants (see Figure 3.4).

3.2 Law of Segregation

- Mendel conducted single-factor crosses in which he followed the variants for a single character (see Figure 3.5).
- The results of his single-factor crosses showed that the dominant trait was always observed in the F_1 generation and displayed a 3:1 ratio in the F_2 generation.
- Based on the results of his single-factor crosses, Mendel proposed three key ideas regarding inheritance. (1) Traits may be dominant or recessive. (2) Genes are passed unaltered from generation to generation. (3) The two copies of a given gene segregate (or separate) from each other during transmission from parent to offspring.
- This third idea is known as the law of segregation (see Figure 3.6).
- A Punnett square can be used to deduce the outcome of crosses.

3.3 Law of Independent Assortment

- By conducting two-factor crosses, Mendel proposed the law of independent assortment, which states that two different genes randomly assort their alleles during the formation of haploid cells (see Figures 3.7–3.9).
- A Punnett square can be used to predict the outcome of two-factor crosses (see Figure 3.10).
- At the molecular level, the recessive variants in Mendel's crosses are due to loss-of-function mutations that prevent the expression of a functionally active protein.

3.4 Chromosome Theory of Inheritance

- The chromosome theory of inheritance describes how the transmission of chromosomes can explain Mendel's laws (see Table 3.1).
- Mendel's law of segregation is explained by the separation of homologs during meiosis (see Figure 3.11).
- Mendel's law of independent assortment is explained by the random alignment of different chromosomes during metaphase of meiosis I (see Figure 3.12).

3.5 Studying Inheritance Patterns in Humans

- Human inheritance patterns are determined by analyzing family trees known as pedigrees (see Figure 3.13).

3.6 Probability and Statistics

- Probability is the number of times an event occurs divided by the total number of events.
- According to the sum rule, the probability that one of two or more mutually exclusive events will occur is equal to the sum of the individual probabilities of the events.
- According to the product rule, the probability of two or more independent events is equal to the product of their individual probabilities. This rule can be used to predict the outcome of crosses involving two or more genes.
- The binomial expansion is used to predict the probability of an unordered combination of events.
- The chi square test is used to test the validity of a hypothesis (see Table 3.2).

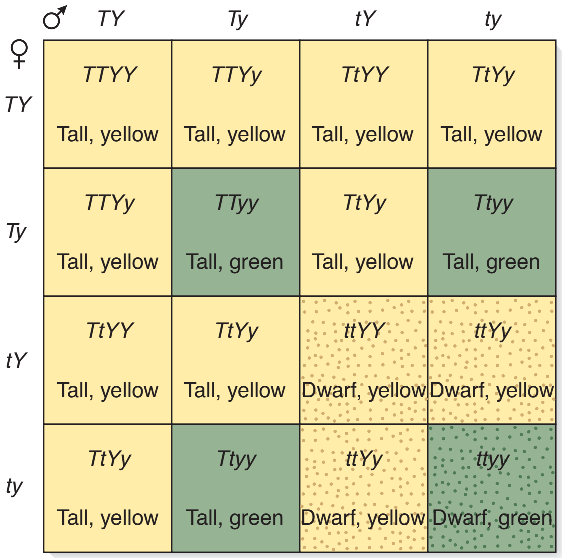
PROBLEM SETS & INSIGHTS

Solved Problems

S1. A heterozygous pea plant that is tall with yellow seeds, *TtYy,* is allowed to self-fertilize. What is the probability that an offspring will be either tall with yellow seeds, tall with green seeds, or dwarf with yellow seeds?

Answer: This problem involves three mutually exclusive events, and so we use the sum rule to solve it. First, we must calculate the individual probabilities for the three phenotypes. The outcome of the cross can be determined using a Punnett square.

Cross: *TtYy* x *TtYy*



♀ ♂	*TY*	*Ty*	*tY*	*ty*
TY	*TTYY* Tall, yellow	*TTYy* Tall, yellow	*TtYY* Tall, yellow	*TtYy* Tall, yellow
Ty	*TTYy* Tall, yellow	*TTyy* Tall, green	*TtYy* Tall, yellow	*Ttyy* Tall, green
tY	*TtYY* Tall, yellow	*TtYy* Tall, yellow	*ttYY* Dwarf, yellow	*ttYy* Dwarf, yellow
ty	*TtYy* Tall, yellow	*Ttyy* Tall, green	*ttYy* Dwarf, yellow	*ttyy* Dwarf, green

$$P_{\text{tall with yellow seeds}} = 9/(9 + 3 + 3 + 1) = 9/16$$

$$P_{\text{tall with green seeds}} = 3/(9 + 3 + 3 + 1) = 3/16$$

$$P_{\text{dwarf with yellow seeds}} = 3/(9 + 3 + 3 + 1) = 3/16$$

$$\text{Sum rule: } 9/16 + 3/16 + 3/16 = 15/16 = 0.94 = 94\%$$

We expect to get one of these three phenotypes 15/16, or 94%, of the time.

S2. For an individual expressing a dominant trait, how can you tell if it is a heterozygote or a homozygote?

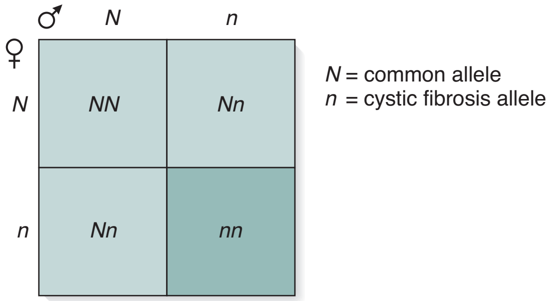
Answer: One way is to conduct a testcross with an individual that expresses the recessive version of the same character. If the individual is heterozygous, half of the offspring will show the recessive trait, but if the individual is homozygous, none of the offspring will express the recessive trait.

$$\begin{array}{ccccc} & Dd \times dd & & \text{or} & DD \times dd \\ & \downarrow & & & \downarrow \\ 1Dd & : & 1\ dd & & \text{All } Dd \\ \text{(dominant trait)} & & \text{(recessive trait)} & & \text{(dominant trait)} \end{array}$$

Another way to determine heterozygosity involves a more careful examination of the individual at the cellular or molecular level. At the cellular level, the heterozygote may not look exactly like the homozygote. This phenomenon is described in Chapter 5. Also, gene cloning methods described in Chapter 19 can be used to distinguish between heterozygotes and homozygotes.

S3. As described in this chapter, a human disease known as cystic fibrosis is inherited as a recessive trait. Two unaffected individuals have a first child with the disease. What is the probability that their next two children will not have the disease?

Answer: An unaffected couple has already produced an affected child. To be affected, the child must be homozygous for the disease allele and thus has inherited one copy from each parent. Therefore, because the parents are unaffected with the disease, we know that both of them must be heterozygous carriers for the recessive disease-causing allele. With this information, we can calculate the probability that they will produce an unaffected offspring. Using a Punnett square, this couple should produce a ratio of 3 unaffected : 1 affected offspring.



♀ ♂	*N*	*n*
N	*NN*	*Nn*
n	*Nn*	*nn*

N = common allele
n = cystic fibrosis allele

The probability of a single unaffected offspring is

$$P_{\text{unaffected}} = 3/(3 + 1) = 3/4$$

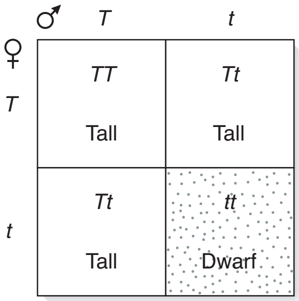
To obtain the probability of getting two unaffected offspring in a row (i.e., in a specified order), we must apply the product rule.

$$3/4 \times 3/4 = 9/16 = 0.56 = 56\%$$

The chance that their next two children will be unaffected is 56%.

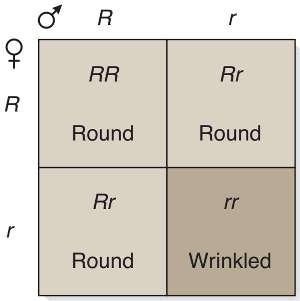
S4. A pea plant is heterozygous for three genes (*Tt Rr Yy*), where T = tall, t = dwarf, R = round seeds, r = wrinkled seeds, Y = yellow seeds, and y = green seeds. If this plant is self-fertilized, what are the predicted phenotypes of the offspring, and what fraction of the offspring will occur in each category?

Answer: You could solve this problem by constructing a large Punnett square and filling in the boxes. However, in this case, eight different male gametes and eight different female gametes are possible: *TRY*, *TRy*, *TrY*, *tRY*, *trY*, *Try*, *tRy*, and *try*. It would become rather tiresome to construct and fill in this Punnett square, which would contain 64 boxes. As an alternative, we can consider each gene separately and then algebraically combine them by multiplying together the expected phenotypic outcomes for each gene. In the cross *Tt Rr Yy* × *Tt Rr Yy*, the following Punnett squares can be made for each gene:



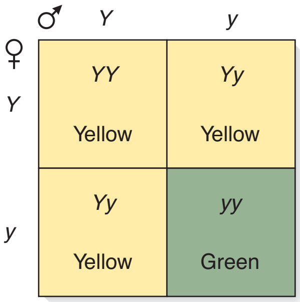
♀ ♂	*T*	*t*
T	*TT* Tall	*Tt* Tall
t	*Tt* Tall	*tt* Dwarf

3 tall : 1 dwarf



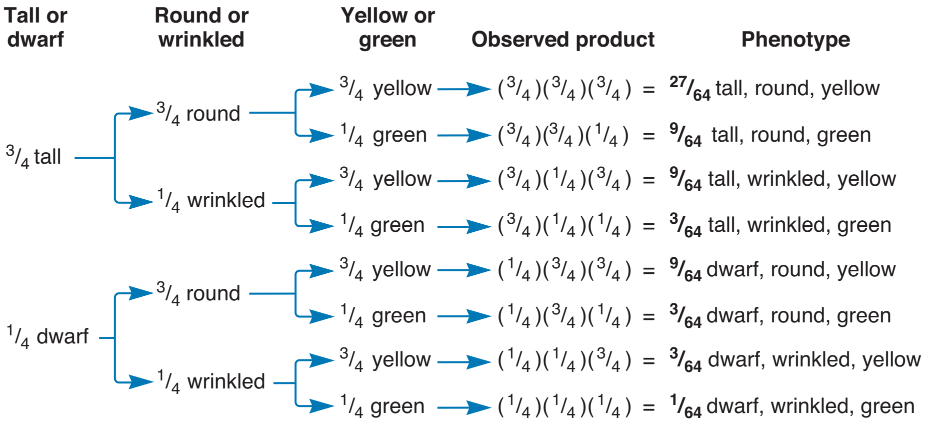
♀ \ ♂	*R*	*r*
R	*RR* Round	*Rr* Round
r	*Rr* Round	*rr* Wrinkled

3 round : 1 wrinkled



♀ \ ♂	*Y*	*y*
Y	*YY* Yellow	*Yy* Yellow
y	*Yy* Yellow	*yy* Green

3 yellow : 1 green



Tall or dwarf	Round or wrinkled	Yellow or green	Observed product	Phenotype
$^3/_4$ tall	$^3/_4$ round	$^3/_4$ yellow	$(^3/_4)(^3/_4)(^3/_4)$ =	$^{27}/_{64}$ tall, round, yellow
		$^1/_4$ green	$(^3/_4)(^3/_4)(^1/_4)$ =	$^9/_{64}$ tall, round, green
	$^1/_4$ wrinkled	$^3/_4$ yellow	$(^3/_4)(^1/_4)(^3/_4)$ =	$^9/_{64}$ tall, wrinkled, yellow
		$^1/_4$ green	$(^3/_4)(^1/_4)(^1/_4)$ =	$^3/_{64}$ tall, wrinkled, green
$^1/_4$ dwarf	$^3/_4$ round	$^3/_4$ yellow	$(^1/_4)(^3/_4)(^3/_4)$ =	$^9/_{64}$ dwarf, round, yellow
		$^1/_4$ green	$(^1/_4)(^3/_4)(^1/_4)$ =	$^3/_{64}$ dwarf, round, green
	$^1/_4$ wrinkled	$^3/_4$ yellow	$(^1/_4)(^1/_4)(^3/_4)$ =	$^3/_{64}$ dwarf, wrinkled, yellow
		$^1/_4$ green	$(^1/_4)(^1/_4)(^1/_4)$ =	$^1/_{64}$ dwarf, wrinkled, green

Instead of constructing a large, 64-box Punnett square, we can use two similar ways to determine the phenotypic outcome of this trihybrid cross. In the multiplication method, we can simply use the product rule and multiply these three combinations together:

(3 tall + 1 dwarf)(3 round + 1 wrinkled)(3 yellow + 1 green)

This multiplication operation can be done in a stepwise manner. First, multiply (3 tall + 1 dwarf) by (3 round + 1 wrinkled).

(3 tall + 1 dwarf)(3 round + 1 wrinkled) = 9 tall, round + 3 tall, wrinkled + 3 dwarf, round + 1 dwarf, wrinkled

Next, multiply this product by (3 yellow + 1 green).

(9 tall, round + 3 tall, wrinkled + 3 dwarf, round + 1 dwarf, wrinkled) (3 yellow + 1 green) = 27 tall, round, yellow + 9 tall, round, green + 9 tall, wrinkled, yellow + 3 tall, wrinkled, green + 9 dwarf, round, yellow + 3 dwarf, round, green + 3 dwarf, wrinkled, yellow + 1 dwarf, wrinkled, green

Even though the multiplication steps are also somewhat tedious, this approach is much easier than making a Punnett square with 64 boxes, filling them in, deducing each phenotype, and then adding them up!

A second approach analogous to the multiplication method is the forked-line method. In this case, the genetic proportions are determined by multiplying together the probabilities of each phenotype. See diagram in the upper right.

S5. In dogs, black fur color is dominant to white. Two heterozygous black dogs are mated. What would be the probability of the following combinations of offspring?

A. A litter of six pups, four with black fur and two with white fur.

B. A litter of six pups, the firstborn with white fur, and among the remaining five pups, two with white fur and three with black fur.

C. A first litter of six pups, four with black fur and two with white fur, and then a second litter of seven pups, five with black fur and two with white fur.

D. A first litter of five pups, four with black fur and one with white fur, and then a second litter of seven pups in which the firstborn is homozygous, the second born is black, and the remaining five pups are three black and two white.

Answer:

A. Because this is an unordered combination of events, we use the binomial expansion equation, where $n = 6$, $x = 4$, $p = 0.75$ (probability of black), and $q = 0.25$ (probability of white).

The answer is 0.297, or 29.7%, of the time.

B. We use the product rule because the order is specified. The first pup is white and then the remaining five are born later. We also need to use the binomial expansion equation to determine the probability of the remaining five pups.

(probability of a white pup)(binomial expansion for the remaining five pups)

The probability of the white pup is 0.25. In the binomial expansion equation, $n = 5$, $x = 2$, $p = 0.25$, and $q = 0.75$.

The answer is 0.066, or 6.6%, of the time.

C. The order of the two litters is specified, so we need to use the product rule. We multiply the probability of the first litter times the probability of the second litter. We need to use the binomial expansion equation for each litter.

(binomial expansion of the first litter)(binomial expansion of the second litter)

For the first litter, $n = 6$, $x = 4$, $p = 0.75$, $q = 0.25$. For the second litter, $n = 7$, $x = 5$, $p = 0.75$, $q = 0.25$.

The answer is 0.092, or 9.2%, of the time.

D. The order of the litters is specified, so we need to use the product rule to multiply the probability of the first litter times the probability of the second litter. We use the binomial expansion equation to determine the probability of the first litter. The probability of the second litter is a little more complicated. The firstborn is homozygous. There are two mutually exclusive ways to be homozygous, *BB* and *bb*. We use the sum rule to determine the probability of the first pup, which equals $0.25 + 0.25 = 0.5$. The probability of the second pup is 0.75, and we use the binomial expansion equation to determine the probability of the remaining pups.

(binomial expansion of first litter)([0.5][0.75][binomial expansion of second litter])

For the first litter, $n = 5$, $x = 4$, $p = 0.75$, $q = 0.25$. For the last five pups in the second litter, $n = 5$, $x = 3$, $p = 0.75$, $q = 0.25$.

The answer is 0.039, or 3.9%, of the time.

S6. In this chapter, the binomial expansion equation is used when only two phenotypic outcomes are possible. When more than two outcomes are possible, we use a multinomial expansion equation to solve a problem involving an unordered number of events. A general expression for this equation is

$$P = \frac{n!}{a!b!c! \ldots} p^a q^b r^c \ldots$$

where P = the probability that the unordered number of events will occur.

$$n = \text{total number of events}$$

$$a + b + c + \ldots = n$$

$$p + q + r + \ldots = 1$$

(p is the likelihood of a, q is the likelihood of b, r is the likelihood of c, and so on)

The multinomial expansion equation can be useful in many genetic problems where more than two combinations of offspring are possible. For example, this formula can be used to solve problems involving an unordered sequence of events in a dihybrid experiment. This approach is illustrated next.

A cross is made between two heterozygous tall plants with axial flowers (*TtAa*), where tall is dominant to dwarf and axial is dominant to terminal flowers. What is the probability that a group of five offspring will be composed of two tall plants with axial flowers, one tall plant with terminal flowers, one dwarf plant with axial flowers, and one dwarf plant with terminal flowers?

Answer:

Step 1. *Calculate the individual probabilities of each phenotype.* This can be accomplished using a Punnett square.

The phenotypic ratios are 9 tall with axial flowers, 3 tall with terminal flowers, 3 dwarf with axial flowers, and 1 dwarf with terminal flowers.

The probability of a tall plant with axial flowers is $9/(9 + 3 + 3 + 1) = 9/16$.

The probability of a tall plant with terminal flowers is $3/(9 + 3 + 3 + 1) = 3/16$.

The probability of a dwarf plant with axial flowers is $3/(9 + 3 + 3 + 1) = 3/16$.

The probability of a dwarf plant with terminal flowers is $1/(9 + 3 + 3 + 1) = 1/16$.

$p = 9/16$

$q = 3/16$

$r = 3/16$

$s = 1/16$

Step 2. *Determine the number of each type of event versus the total number of events.*

$n = 5$

$a = 2$

$b = 1$

$c = 1$

$d = 1$

Step 3. *Substitute the values in the multinomial expansion equation.*

$$P = \frac{n!}{a!b!c!d!} p^a q^b r^c s^d$$

$$P = \frac{5!}{2!1!1!1!} (9/16)^2 (3/16)^1 (3/16)^1 (1/16)^1$$

$$P = 0.04 = 4\%$$

This means that 4% of the time we would expect to obtain five offspring with the phenotypes described in the question.

Conceptual Questions

C1. Why did Mendel's work refute the idea of blending inheritance?

C2. What is the difference between cross-fertilization and self-fertilization?

C3. Describe the difference between genotype and phenotype. Give three examples. Is it possible for two individuals to have the same phenotype but different genotypes?

C4. With regard to genotypes, what is a true-breeding organism?

C5. How can you determine whether an organism is heterozygous or homozygous for a dominant trait?

C6. In your own words, describe what Mendel's law of segregation means. Do not use the word *segregation* in your answer.

C7. Based on genes in pea plants that we have considered in this chapter, which statement(s) is incorrect?

A. The gene causing tall plants is an allele of the gene causing dwarf plants.

B. The gene causing tall plants is an allele of the gene causing purple flowers.

C. The alleles causing tall plants and purple flowers are dominant.

C8. In a cross between a heterozygous tall pea plant and a dwarf plant, predict the ratios of the offspring's genotypes and phenotypes.

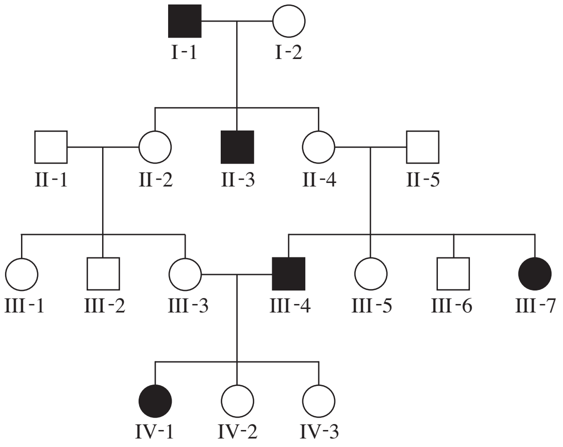
C9. Do you know the genotype of an individual with a recessive trait or a dominant trait? Explain your answer.

C10. A cross is made between a pea plant that has constricted pods (a recessive trait; smooth is dominant) and is heterozygous for seed color (yellow is dominant to green) and a plant that is heterozygous for both pod texture and seed color. Construct a Punnett square that depicts this cross. What are the predicted outcomes of genotypes and phenotypes of the offspring?

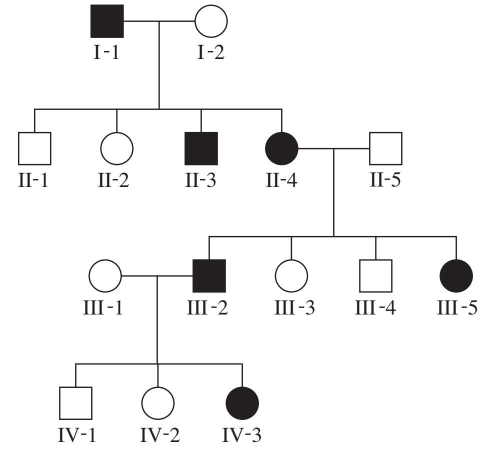
C11. A pea plant that is heterozygous with regard to seed color (yellow is dominant to green) is allowed to self-fertilize. What are the predicted outcomes of genotypes and phenotypes of the offspring?

C12. Describe the significance of nonparentals with regard to the law of independent assortment. In other words, explain how the appearance of nonparentals refutes a linkage hypothesis.

C13. For the following two pedigrees, describe what you think is the most likely inheritance pattern (dominant versus recessive). Explain your reasoning. Filled (black) symbols indicate affected individuals.



(a)



(b)

C14. Ectrodactyly, also known as "lobster claw syndrome," is a recessive disorder in humans. If a phenotypically unaffected couple produces an affected offspring, what are the following probabilities?

A. Both parents are heterozygotes.

B. An offspring is a heterozygote.

C. The next three offspring will be phenotypically unaffected.

D. Any two out of the next three offspring will be phenotypically unaffected.

C15. Identical twins are produced from the same sperm and egg (which splits after the first mitotic division), whereas fraternal twins are produced from separate sperm and separate egg cells. If two parents with brown eyes (a dominant trait) produce one twin boy with blue eyes, what are the following probabilities?

A. If the other twin is identical, he will have blue eyes.

B. If the other twin is fraternal, he or she will have blue eyes.

C. If the other twin is fraternal, he or she will transmit the blue eye allele to his or her offspring.

D. The parents are both heterozygotes.

C16. In cocker spaniels, solid coat color is dominant over spotted coat color. If two heterozygous dogs were crossed to each other, what would be the probability of the following combinations of offspring?

A. A litter of five pups, four with solid fur and one with spotted fur.

B. A first litter of six pups, four with solid fur and two with spotted fur, and then a second litter of five pups, all with solid fur.

C. A first litter of five pups, the firstborn with solid fur, and then among the next four, three with solid fur and one with spotted fur, and then a second litter of seven pups in which the firstborn is spotted, the second born is spotted, and the remaining five are composed of four solid and one spotted animal.

D. A litter of six pups, the firstborn with solid fur, the second born spotted, and among the remaining four pups, two with spotted fur and two with solid fur.

C17. A cross was made between a white male dog and two different black females. The first female gave birth to eight black pups, and the second female gave birth to four white and three black pups. What are the likely genotypes of the male parent and the two female parents? Explain whether you are uncertain about any of the genotypes.

C18. In humans, the allele for brown eye color (*B*) is dominant to blue eye color (*b*). If two heterozygous parents produce children, what are the following probabilities?

A. The first two children have blue eyes.

B. A total of four children, two with blue eyes and the other two with brown eyes.

C. The first child has blue eyes, and the next two have brown eyes.

C19. Albinism, a condition characterized by a partial or total lack of skin pigment, is a recessive human trait. If a phenotypically unaffected couple produced an albino child, what is the probability that their next child will be albino?

C20. A true-breeding tall plant was crossed to a dwarf plant. Tallness is a dominant trait. The F_1 individuals were allowed to self-fertilize. What are the following probabilities for the F_2 generation?

A. The first plant is dwarf.

B. The first plant is dwarf or tall.

C. The first three plants are tall.

D. For any seven plants, three are tall and four are dwarf.

E. The first plant is tall, and then among the next four, two are tall and the other two are dwarf.

C21. For pea plants with the following genotypes, list the possible gametes that the plant can make:

A. *TT Yy Rr*

B. *Tt YY rr*

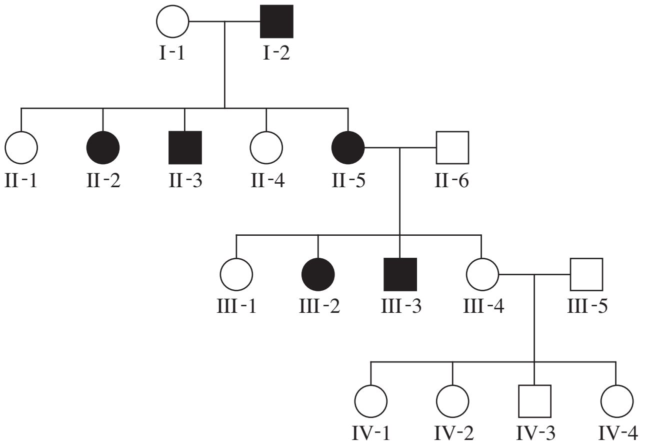
C. *Tt Yy Rr*

D. *tt Yy rr*

C22. An individual has the genotype *Aa Bb Cc* and makes an abnormal gamete with the genotype *AaBc*. Does this gamete violate the law of independent assortment or the law of segregation (or both)? Explain your answer.

C23. In people with maple syrup urine disease, the body is unable to metabolize the amino acids leucine, isoleucine, and valine. One of the symptoms is that the urine smells like maple syrup. An unaffected couple produced six children in the following order: unaffected daughter, affected daughter, unaffected son, unaffected son, affected son, and unaffected son. The youngest unaffected son marries an unaffected woman and has three children in the following order: affected daughter, unaffected daughter, and unaffected son. Draw a pedigree that describes this family. What type of inheritance (dominant or recessive) would you propose to explain maple syrup urine disease?

C24. Marfan syndrome is a rare inherited human disorder characterized by unusually long limbs and digits plus defects in the heart (especially the aorta) and the eyes, among other symptoms. Following is a pedigree for this disorder. Affected individuals are shown with filled (black) symbols. What type of inheritance pattern do you think is the most likely?



C25. A true-breeding pea plant with round and green seeds was crossed to a true-breeding plant with wrinkled and yellow seeds. Round and yellow seeds are the dominant traits. The F_1 plants were allowed to self-fertilize. What are the following probabilities for the F_2 generation?

A. An F_2 plant with wrinkled, yellow seeds.

B. Three out of three F_2 plants with round, yellow seeds.

C. Five F_2 plants in the following order: two have round, yellow seeds; one has round, green seeds; and two have wrinkled, green seeds.

D. An F_2 plant will not have round, yellow seeds.

C26. A true-breeding tall pea plant was crossed to a true-breeding dwarf plant. What is the probability that an F_1 individual will be true-breeding? What is the probability that an F_1 individual will be a true-breeding tall plant?

C27. What are the expected phenotypic ratios from the following cross: *Tt Rr yy Aa* × *Tt rr YY Aa*, where *T* = tall, *t* = dwarf, *R* = round, *r* = wrinkled, *Y* = yellow, *y* = green, *A* = axial, *a* = terminal; *T*, *R*, *Y*, and *A* are dominant alleles. Note: See Solved Problem S4 for help in answering this problem.

C28. When an abnormal organism contains three copies of a gene (instead of the normal number of two copies), the alleles for the gene usually segregate so that a gamete contains one or two copies of the gene. Let's suppose that an abnormal pea plant has three copies of the height gene. Its genotype is *TTt*. The plant is also heterozygous for the seed color gene, *Yy*. How many types of gametes can this plant make, and in what proportions? (Assume that it is equally likely that a gamete will contain one or two copies of the height gene.)

C29. Honeybees are unusual in that male bees (drones) have only one copy of each gene, but female bees have two copies of their genes. That is because drones develop from eggs that have not been fertilized by sperm cells. In bees, the trait of long wings is dominant over short wings, and the trait of black eyes is dominant over white eyes. If a drone with short wings and black eyes was mated to a queen bee that is heterozygous for both genes, what are the predicted genotypes and phenotypes of male and female offspring? What are the phenotypic ratios if we assume an equal number of male and female offspring?

C30. A pea plant that is dwarf with green, wrinkled seeds was crossed to a true-breeding plant that is tall with yellow, round seeds. The F_1 generation was allowed to self-fertilize. What types of gametes, and in what proportions, would the F_1 generation make? What would be the ratios of genotypes and phenotypes of the F_2 generation?

C31. A true-breeding plant with round and green seeds was crossed to a true-breeding plant with wrinkled and yellow seeds. The F_1 plants were allowed to self-fertilize. What is the probability of obtaining the following plants in the F_2 generation: two that have round, yellow seeds; one with round, green seeds; and two with wrinkled, green seeds? (*Note:* See Solved Problem S6 for help.)

C32. Woolly hair is a rare dominant trait found in people of Scandinavian descent in which the hair resembles the wool of a sheep. A male with woolly hair, who has a mother with straight hair, moves to an island that is inhabited by people who are not of Scandinavian descent. Assuming that no other Scandinavians immigrate to the island, what is the probability that a great-grandchild of this male will have woolly hair? (*Hint:* You may want to draw a pedigree to help you figure this out.) If this woolly-haired male has eight great-grandchildren, what is the probability that one out of eight will have woolly hair?

C33. Huntington disease is a rare dominant trait that causes neurodegeneration later in life. A man in his thirties, who already has three children, discovers that his mother has Huntington disease though his father is unaffected. What are the following probabilities?

A. That the man in his thirties will develop Huntington disease.

B. That his first child will develop Huntington disease.

C. That one out of three of his children will develop Huntington disease.

C34. A woman with achondroplasia (a dominant form of dwarfism) and a phenotypically unaffected man have seven children, all of whom have achondroplasia. What is the probability of producing such a family if this woman is a heterozygote? What is the probability that the woman is a heterozygote if her eighth child does not have this disorder?

Application and Experimental Questions

E1. Describe three advantages of using pea plants as an experimental organism.

E2. Explain the technical differences between a cross-fertilization experiment versus a self-fertilization experiment.

E3. How long did it take Mendel to complete the experiment in Figure 3.5?

E4. For all seven characters described in the data of Figure 3.5, Mendel allowed the F_2 plants to self-fertilize. He found that when F_2 plants with recessive traits were crossed to each other, they always bred true. However, when F_2 plants with dominant traits were crossed, some bred true but others did not. A summary of Mendel's results is shown here.

The Ratio of True-Breeding and Non-True-Breeding Parents of the F_2 Generation

F_2 Parents	True-Breeding	Non-True-Breeding	Ratio
Round	193	372	1:1.93
Yellow	166	353	1:2.13
Gray	36	64	1:1.78
Smooth	29	71	1:2.45
Green	40	60	1:1.5
Axial	33	67	1:2.08
Tall	28	72	1:2.57
TOTAL:	525	1059	1:2.02

When considering the data in this table, keep in mind that it describes the characteristics of the F_2 generation parents that had displayed a dominant phenotype. These data were deduced by analyzing the outcome of the F_3 generation. Based on Mendel's laws, explain the 1:2 ratio obtained in these data.

E5. From the point of view of crosses and data collection, what are the experimental differences between a single-factor and a two-factor experiment?

E6. As in many animals, albino coat color is a recessive trait in guinea pigs. Researchers removed the ovaries from an albino female guinea pig and then transplanted ovaries from a true-breeding black guinea pig. They then mated this albino female (with the transplanted ovaries) to an albino male. The albino female produced three offspring. What were their coat colors? Explain the results.

E7. The fungus *Melampsora lini* causes a disease known as flax rust. Different strains of *M. lini* cause varying degrees of the rust disease. Conversely, different strains of flax are resistant or sensitive to the various varieties of rust. The Bombay variety of flax is resistant to *M. lini*-strain 22 but sensitive to *M. lini*-strain 24. A strain of flax called 770B is just the opposite; it is resistant to strain 24 but sensitive to strain 22. When 770B was crossed to Bombay, all F_1 individuals were resistant to both strain 22 and strain 24. When F_1 individuals were self-fertilized, the following data were obtained:

43 resistant to strain 22 but sensitive to strain 24

9 sensitive to strain 22 and strain 24

32 sensitive to strain 22 but resistant to strain 24

110 resistant to strain 22 and strain 24

Explain the inheritance pattern for flax resistance and sensitivity to *M. lini* strains.

E8. For Mendel's data shown in Figure 3.8, conduct a chi square analysis to determine if the data agree with Mendel's law of independent assortment.

E9. Would it be possible to deduce the law of independent assortment from a single-factor experiment? Explain your answer.

E10. In fruit flies, curved wings are recessive to straight wings, and ebony body is recessive to gray body. A cross was made between true-breeding flies with curved wings and gray bodies to flies with straight wings and ebony bodies. The F_1 offspring were then mated to flies with curved wings and ebony bodies to produce an F_2 generation.

A. Diagram the genotypes of this cross, starting with the parental generation and ending with the F_2 generation.

B. What are the predicted phenotypic ratios of the F_2 generation?

C. Let's suppose the following data were obtained for the F_2 generation:

114 curved wings, ebony body

105 curved wings, gray body

111 straight wings, gray body

114 straight wings, ebony body

Conduct a chi square analysis to determine if the experimental data are consistent with the expected outcome based on Mendel's laws.

E11. A recessive allele in mice results in an abnormally long neck. Sometimes, during early embryonic development, the abnormal neck causes the embryo to die. An experimenter began with a population of true-breeding normal mice and true-breeding mice with long necks. Crosses were made between these two populations to produce an F_1 generation of mice with normal necks. The F_1 mice were then mated to each other to obtain an F_2 generation. For the mice that were born alive, the following data were obtained:

522 mice with normal necks

62 mice with long necks

What percentage of homozygous mice (that would have had long necks if they had survived) died during embryonic development?

E12. The data in Figure 3.5 show the results of the F_2 generation for seven of Mendel's crosses. Conduct a chi square analysis to determine if these data are consistent with the law of segregation.

E13. Let's suppose you conducted an experiment involving genetic crosses and calculated a chi square value of 1.005. There were four categories of offspring (i.e., the degrees of freedom equaled 3). Explain what the 1.005 value means. Your answer should include the phrase "80% of the time."

E14. A tall pea plant with axial flowers was crossed to a dwarf plant with terminal flowers. Tall plants and axial flowers are dominant traits. The following offspring were obtained: 27 tall, axial flowers; 23 tall, terminal flowers; 28 dwarf, axial flowers; and 25 dwarf, terminal flowers. What are the genotypes of the parents?

E15. A cross was made between two strains of plants that are agriculturally important. One strain was disease-resistant but herbicide-sensitive; the other strain was disease-sensitive but herbicide-resistant. A plant breeder crossed the two plants and then allowed the F_1 generation to self-fertilize. The following data were obtained:

F_1 generation: All offspring are disease-sensitive and herbicide-resistant

F_2 generation: 157 disease-sensitive, herbicide-resistant
57 disease-sensitive, herbicide-sensitive
54 disease-resistant, herbicide-resistant
20 disease-resistant, herbicide-sensitive

Total: 288

Formulate a hypothesis that you think is consistent with the observed data. Test the goodness of fit between the data and your hypothesis using a chi square test. Explain what the chi square results mean.

E16. A cross was made between a plant that has blue flowers and purple seeds to a plant with white flowers and green seeds. The following data were obtained:

F_1 generation: All offspring have blue flowers with purple seeds

F_2 generation: 103 blue flowers, purple seeds
49 blue flowers, green seeds
44 white flowers, purple seeds
104 white flowers, green seeds

Total: 300

Start with the hypothesis that blue flowers and purple seeds are dominant traits and that the two genes assort independently. Calculate a chi square value. What does this value mean with regard to your hypothesis? If you decide to reject your hypothesis, which aspect of the hypothesis do you think is incorrect (i.e., blue flowers and purple seeds are dominant traits, or the idea that the two genes assort independently)?

E17. Discuss why crosses (i.e., the experiments of Mendel) and the microscopic observations of chromosomes during mitosis and meiosis were both needed to deduce the chromosome theory of inheritance.

Questions for Student Discussion/Collaboration

1. Consider this cross in pea plants: $Tt\ Rr\ yy\ Aa \times Tt\ rr\ Yy\ Aa$, where T = tall, t = dwarf, R = round, r = wrinkled, Y = yellow, y = green, A = axial, a = terminal. What is the expected phenotypic outcome of this cross? Have one group of students solve this problem by making one big Punnett square, and have another group solve it by making four single-gene Punnett squares and using the product rule. Time each other to see who gets done first.
2. A cross was made between two pea plants, $TtAa$ and $Ttaa$, where T = tall, t = dwarf, A = axial, and a = terminal. What is the probability that the first three offspring will be tall with axial flowers or dwarf with terminal flowers and the fourth offspring will be tall with axial flowers. Discuss what operation(s) (e.g., sum rule, product rule, or binomial expansion equation) you used to solve them and in what order they were used.
3. Consider this four-factor cross: $Tt\ Rr\ yy\ Aa \times Tt\ RR\ Yy\ aa$, where T = tall, t = dwarf, R = round, r = wrinkled, Y = yellow, y = green, A = axial, a = terminal. What is the probability that the first three plants will have round seeds? What is the easiest way to solve this problem?

Answers to Comprehension Questions

3.1: d, a, b

3.2: c, b, b

3.3: b, c, a

3.4: c, a

3.5: b, c

3.6: a, b (use the binomial expansion), d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

4

Opposite sexes. *Most species of animals, such as these male and female cardinals, are found in two sexes.*

CHAPTER OUTLINE

4.1 Mechanisms of Sex Determination Among Various Species

4.2 Dosage Compensation and X Inactivation in Mammals

4.3 Properties of the X and Y Chromosome in Mammals

4.4 Transmission Patterns for X-linked Genes

SEX DETERMINATION AND SEX CHROMOSOMES

In Chapter 2, we examined the process of sexual reproduction in which two gametes fuse with each other to begin the life of a new individual. Within a population, sexual reproduction enhances genetic diversity because the genetic material of offspring comes from two sources. For most species of animals and some species of plants, sexual reproduction is carried out by individuals of the opposite sex—females and males. The underlying mechanism by which an individual develops into a female or a male is called **sex determination.** As we will see, a variety of mechanisms promote this process.

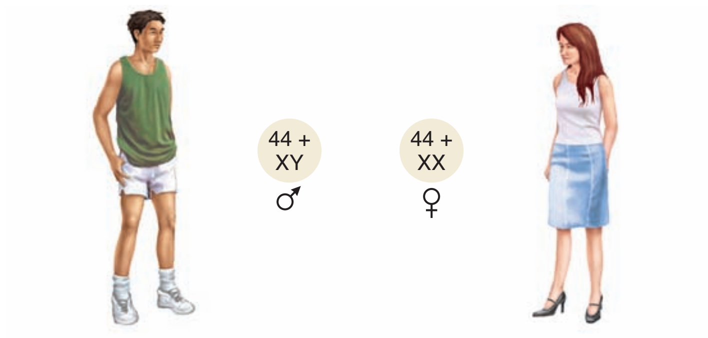
For some species, females and males differ in their genomes. For example, you are probably already familiar with the idea that people differ with regard to X and Y chromosomes. Females are XX and males are XY, which means that females have two copies of the X chromosome, whereas males have one X and one Y chromosome. Because these chromosomes carry different genes, chromosomal differences between the sexes also result in unique phenotypes and inheritance patterns that differ from those we discussed in Chapter 3. In this chapter, we will explore how genes located on the X chromosome exhibit a characteristic pattern of inheritance.

4.1 MECHANISMS OF SEX DETERMINATION AMONG VARIOUS SPECIES

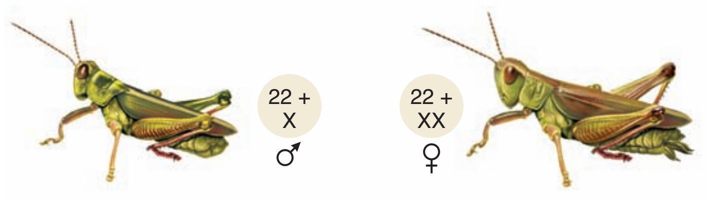
After gametes fuse with each other during fertilization, what factor(s) determine whether the resulting zygote and embryo develop into a female or a male? Researchers have studied the process of sex determination in a wide range of species and discovered that several different mechanisms exist. In this section, we will explore some common mechanisms of sex determination in animals and plants.

Sex Differences May Depend on the Presence of Sex Chromosomes

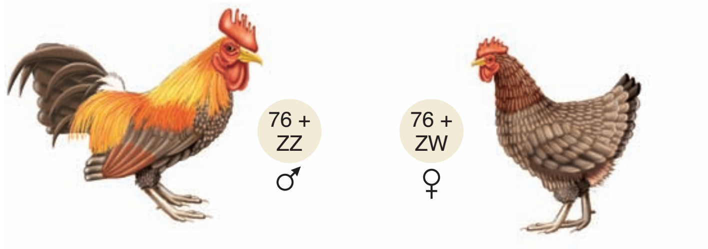
According to the chromosome theory of inheritance, which we discussed in Chapter 3, chromosomes carry the genes that determine an organism's traits. Not surprisingly, sex determination in



(a) X–Y system in mammals



(b) The X–0 system in certain insects



(c) The Z–W system in birds

FIGURE 4.1 **Sex determination via the presence of sex chromosomes.** See text for a description.

Genes→Traits Certain genes that are found on the sex chromosomes play a key role in the development of sex (male vs. female). For example, in mammals, a gene on the Y chromosome initiates male development. In the X-0 system, the ratio of X chromosomes to the sets of autosomes plays a key role in governing the pathway of development toward male or female.

Concept check: *What is the difference between the X-Y and X-0 systems of sex determination?*

some species is due to the presence of particular chromosomes. In 1901, Clarence McClung, who studied grasshoppers, was the first to suggest that male and female sexes are due to the inheritance of particular chromosomes. Since McClung's initial observations, we now know that a pair of chromosomes, called the **sex chromosomes,** determines sex in many different species. Some examples are described in **Figure 4.1**.

In the X-Y system of sex determination, which operates in mammals, the male contains one X chromosome and one Y chromosome, whereas the female contains two X chromosomes (**Figure 4.1a**). In this case, the male is called the **heterogametic sex.** Two types of sperm are produced: one that carries only the X chromosome, and another that carries the Y. In contrast, the female is the **homogametic sex** because all eggs carry a single X chromosome. The 46 chromosomes carried by humans consist of 1 pair of sex chromosomes and 22 pairs of **autosomes**—chromosomes that are not sex chromosomes. In the human male, each of the four sperm produced during gametogenesis contains 23 chromosomes. Two sperm contain an X chromosome, and the other two have a Y chromosome. The sex of the offspring is determined by whether the sperm that fertilizes the egg carries an X or a Y chromosome.

What causes an offspring to develop into a male or female? One possibility is that two X chromosomes are required for female development. A second possibility is that the Y chromosome promotes male development. In the case of mammals, the second possibility is correct. This is known from the analysis of rare individuals who carry chromosomal abnormalities. For example, mistakes that occasionally occur during meiosis may produce an individual who carries two X chromosomes and one Y chromosome. Such an individual develops into a male. In addition, people are sometimes born with a single X chromosome and not another sex chromosome. Such individuals become females. The chromosomal basis for sex determination in mammals is rooted in the location of a particular gene on the Y chromosome. The presence of a gene on the Y chromosome called the *Sry* gene causes maleness.

Another mechanism of sex determination that involves sex chromosomes is the X-0 system that operates in many insects (**Figure 4.1b**). In some insect species, the male has only one sex chromosome (the X) and is designated X0, whereas the female has a pair (two X's). In other insect species, such as *Drosophila melanogaster*, the male is XY. For both types of insect species (i.e., X0 or XY males, and XX females), the ratio between X chromosomes and the number of autosomal sets determines sex. If a fly has one X chromosome and is diploid for the autosomes ($2n$), the ratio is 1/2, or 0.5. This fly becomes a male even if it does not receive a Y chromosome. In contrast to mammals, the Y chromosome in the X-0 system does not determine maleness. If a fly receives two X chromosomes and is diploid, the ratio is 2/2, or 1.0, and the fly becomes a female.

For the Z-W system, which determines sex in birds and some fish, the male is ZZ and the female is ZW (**Figure 4.1c**). The letters Z and W are used to distinguish these types of sex chromosomes from those found in the X-Y pattern of sex determination of other species. In the Z-W system, the male is the homogametic sex, and the female is heterogametic.

Sex Differences May Depend on the Number of Sets of Chromosomes

Another interesting mechanism of sex determination, known as the **haplodiploid system,** is found in bees, wasps, and ants (**Figure 4.2**). For example, in honey bees, the male, which is called a drone, is produced from unfertilized haploid eggs. Male honeybees contain a single set of 16 chromosomes. By comparison, female honeybees, both worker bees and queen bees, are produced from fertilized eggs and therefore are diploid. They contain two sets of chromosomes, for a total of 32. In this case, only females are produced by sexual reproduction.

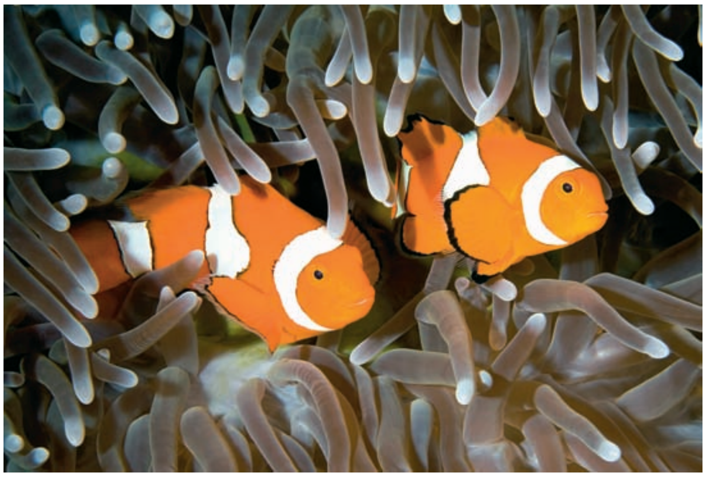
Male honeybee (Drone)
Haploid – 16 chromosomes

Female honeybee
Diploid – 32 chromosomes

FIGURE 4.2 The haplodiploid mechanism of sex determination. In this system, males are haploid, whereas females are diploid.

Concept check: Is the male bee produced by sexual reproduction? Explain.

(a) Sex determination via temperature.
American alligator (*A. mississippiensis*)



(b) Sex determination via behavior.
Clownfish (*Amphiprion ocellaris*)

FIGURE 4.3 Sex determination caused by environmental factors. (**a**) In the alligator, temperature determines whether an individual develops into a female or male. (**b**) In clownfish, males can change into females due to behavioral changes that occur when a dominant female dies.

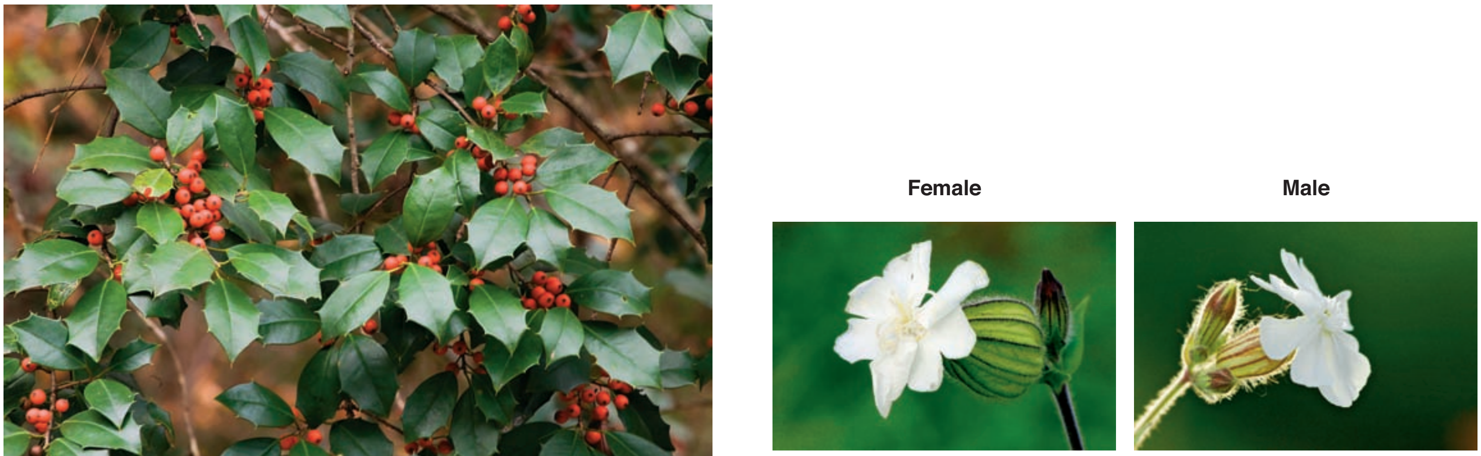
Concept check: How might global warming affect alligator populations?

Sex Differences May Depend on the Environment

Although sex in many species of animals is determined by chromosomes, other mechanisms are also known. In certain reptiles and fish, sex is controlled by environmental factors such as temperature. For example, in the American alligator (*Alligator mississippiensis*), temperature controls sex development (**Figure 4.3a**). When fertilized eggs of this alligator species are incubated at 33°C, nearly 100% of them produce male individuals. In contrast, eggs incubated at a temperature a few degrees below 33°C produce nearly all females, whereas those incubated a few degrees above 33°C produce about 95% females.

Another way that sex can be environmentally determined is via behavior. Clownfish of the genus *Amphiprion* are coral reef fish that live among anemones on the ocean floor (**Figure 4.3b**). One anemone typically harbors a harem of clownfish consisting of a large female, a medium-sized reproductive male, and small nonreproductive juveniles. Clownfish are **protandrous hermaphrodites**—they can switch from male to female! When the female of a harem dies, the reproductive male changes sex to become a female and the largest of the juveniles matures into a reproductive male. Unlike male and female humans, the opposite sexes of clownfish are not determined by chromosome differences. Male and female clownfish have the same chromosomal composition.

How can a clownfish switch from female to male? A juvenile clownfish has both male and female immature sexual organs. Hormone levels, particularly those of an androgen called



(a) American holly (*I. opaca*)

(b) Female and male flowers on separate individuals in white campion (*S. latifolia*)

FIGURE 4.4 **Examples of dioecious plants in which individuals produce only male gametophytes or only female gametophytes.** (**a**) American holly (*Ilex opaca*). The female sporophyte, which produces red berries, is shown here. (**b**) White campion (*Silene latifolia*), which is often studied by researchers.

Concept check: Which are the opposite sexes in dioecious plants—the sporophytes or the gametophytes?

testosterone and an estrogen called estradiol, control the expression of particular genes. In nature, the first sexual change that usually happens is that a juvenile clownfish becomes a male. This occurs when the testosterone level becomes high, which promotes the expression of genes that encode proteins that cause the male organs to mature. Later, when the female of the harem dies, the estradiol level in the reproductive male becomes high and testosterone is decreased. This alters gene expression in a way that leads to the synthesis of some new types of proteins and prevents the synthesis of others. In other words, changes in hormones alter the composition of the proteome. When this occurs, the female organs grow and the male reproductive system degenerates. The male fish becomes female. These sex changes, which are irreversible, are due to sequential changes in the proteomes of clownfish.

What factor determines the hormone levels in clownfish? A female seems to control the other clownfish in the harem through aggressive dominance, thereby preventing the formation of other females. This aggressive behavior suppresses an area of the brain in the other clownfish that is responsible for the production of certain hormones that are needed to promote female development. If a clownfish is left by itself in an aquarium, it will automatically develop into a female because this suppression does not occur.

Dioecious Plant Species Have Opposite Sexes

In most plant species, a single diploid individual (a sporphyte) produces both female and male gametophytes, which are haploid and contain egg or sperm cells, respectively (see Chapter 2, Figure 2.15). However, some plant species are **dioecious,** which means that some individuals produce only male gametophytes, whereas others produce only female gametophytes. These include hollies (**Figure 4.4a**), willows, and ginkgo trees. The genetics of sex determination in dioecious plant species is beginning to emerge. To study this process, many researchers have focused their attention on the white campion, *Silene latifolia*, which is a relatively small dioecious plant with a short generation time (**Figure 4.4b**). In this species, sex chromosomes, designated X and Y, are responsible for sex determination. The male plant has X and Y chromosomes, whereas the female plant is XX. Sex chromosomes are also found in other plant species such as papaya and spinach. However, in other dioecious species, cytological examination of the chromosomes does not always reveal distinct types of sex chromosomes. Even so, in these plant species, the male plants usually appear to be the heterogametic sex.

4.1 REVIEWING THE KEY CONCEPTS

- In the X-Y, X-0, and Z-W systems, sex is determined by the presence and number of particular sex chromosomes (see Figure 4.1).
- In the haplodiploid system, sex is determined by the number of sets of chromosomes (see Figure 4.2).
- In some species, such as alligators and clownfish, sex is determined by environmental factors (see Figure 4.3).
- Dioecious plants exist as individuals that produce only pollen and those that produce only eggs. Sex chromosomes sometimes determine sex in these species (see Figure 4.4).

4.1 COMPREHENSION QUESTIONS

1. Among different species, sex may be determined by
   a. differences in sex chromosomes.
   b. differences in the number of sets of chromosomes.
   c. environmental factors.
   d. all of the above.

2. In mammals, sex is determined by
   a. the *Sry* gene on the Y chromosome.
   b. having two copies of the X chromosome.
   c. having one copy of the X chromosome.
   d. both a and c.
3. An abnormal fruit fly has two sets of autosomes and is XXY. Such a fly would be
   a. a male.
   b. a female.
   c. a hermaphrodite
   d. none of the above.

4.2 DOSAGE COMPENSATION AND X INACTIVATION IN MAMMALS

Dosage compensation refers to the phenomenon in which the level of expression of many genes on the sex chromosomes (e.g., the X chromosome) is similar in both sexes, even though males and females have a different complement of sex chromosomes. This term was coined in 1932 by Hermann Muller to explain the effects of eye color mutations in *Drosophila*. Muller observed that female flies homozygous for certain alleles on the X chromosome had a similar phenotype to males, which have only one copy of the gene. He noted that an allele on the X chromosome conferring an apricot eye color produces a very similar phenotype in a female carrying two copies of the gene and in a male with just one. In contrast, a female that has one copy of the apricot allele and a deletion of the apricot allele on the other X chromosome has eyes of paler color. Therefore, one copy of the allele in the female is not equivalent to one copy of the allele in the male. Instead, two copies of the allele in the female produce a phenotype that is similar to that produced by one copy in the male. In other words, the difference in gene dosage—two copies in females versus one copy in males—is being compensated for at the level of gene expression. In this section, we will explore how this occurs in different species of animals.

Dosage Compensation Is Necessary in Some Species to Ensure Genetic Equality Between the Sexes

Dosage compensation has been studied extensively in mammals, *Drosophila*, and *Caenorhabditis elegans* (a nematode). Depending on the species, dosage compensation occurs via different mechanisms (**Table 4.1**). Female mammals equalize the expression of genes on the X chromosome by turning off one of their two X chromosomes. This process is known as **X inactivation.** In *Drosophila*, the male accomplishes dosage compensation by doubling the expression of most genes on the X chromosome. In *C. elegans*, the XX animal is a hermaphrodite that produces both sperm and egg cells, and an animal carrying a single X chromosome is a male that produces only sperm. The XX hermaphrodite diminishes the expression of genes on each X chromosome to approximately 50%.

In birds, the Z chromosome is a large chromosome, usually the fourth or fifth largest, and contains many genes. The W chromosome is generally a much smaller microchromosome containing a high proportion of repeat-sequence DNA that does not encode genes. Males are ZZ and females are ZW. Several years ago, researchers studied the level of expression of a Z-linked gene that encodes an enzyme called aconitase. They discovered that males express twice as much aconitase as females do. These results suggested that dosage compensation does not occur in birds. More recently, the expression of hundreds of Z-linked genes has been examined in chickens. These newer results also suggest that birds lack a general mechanism of dosage compensation that controls the expression of most Z-linked genes. Even so, the pattern of gene expression between males and females was found to vary a great deal for certain Z-linked genes. Overall, the results suggest that some Z-linked genes may be dosage-compensated, but many of them are not.

Dosage Compensation Occurs in Female Mammals by the Inactivation of One X Chromosome

In 1961, Mary Lyon proposed that dosage compensation in mammals occurs by the inactivation of a single X chromosome in

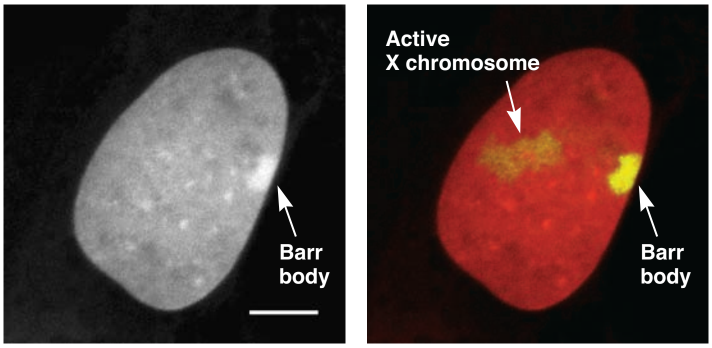
TABLE **4.1**

Mechanisms of Dosage Compensation Among Different Species

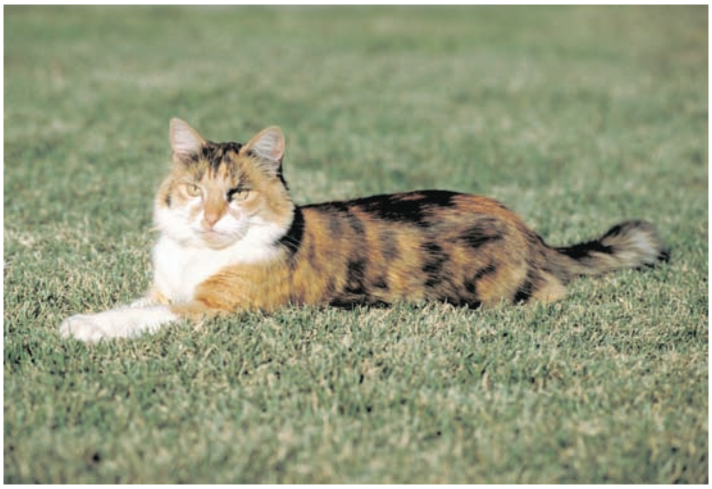
Species	Sex Chromosomes in: Females	Sex Chromosomes in: Males	Mechanism of Compensation
Placental mammals	XX	XY	One of the X chromosomes in the somatic cells of females is inactivated. In certain species, the paternal X chromosome is inactivated, and in other species, such as humans, either the maternal or paternal X chromosome is randomly inactivated throughout the somatic cells of females.
Marsupial mammals	XX	XY	The paternally derived X chromosome is inactivated in the somatic cells of females.
Drosophila melanogaster	XX	XY	The level of expression of genes on the X chromosome in males is increased twofold.
Caenorhabditis elegans	XX*	X0	The XX hermaphrodite diminishes the expression of genes on each X chromosome to about 50%.

*In *C. elegans*, an XX individual is a hermaphrodite, not a female.

females. Liane Russell also proposed the same idea around the same time. This proposal brought together two lines of study. The first type of evidence came from cytological studies. In 1949, Murray Barr and Ewart Bertram identified a highly condensed structure in the interphase nuclei of somatic cells in female cats that was not found in male cats. This structure became known as the **Barr body** (**Figure 4.5a**). In 1960, Susumu Ohno correctly proposed that the Barr body is a highly condensed X chromosome.



(a) Nucleus with a Barr body



(b) A calico cat

FIGURE 4.5 **X chromosome inactivation in female mammals.** **(a)** The left micrograph shows the Barr body on the periphery of a human nucleus after staining with a DNA-specific dye. Because it is compact, the Barr body is the most brightly staining. The white scale bar is 5 μm. The right micrograph shows the same nucleus using a yellow fluorescent probe that recognizes the X chromosome. The Barr body is more compact than the active X chromosome, which is to the left of the Barr body. **(b)** The fur pattern of a calico cat.

Genes→Traits The pattern of black and orange fur on this cat is due to random X inactivation during embryonic development. The orange patches of fur are due to the inactivation of the X chromosome that carries a black allele; the black patches are due to the inactivation of the X chromosome that carries the orange allele. In general, only heterozygous female cats can be calico. A rare exception is a male cat (XXY) that has an abnormal composition of sex chromosomes.

Concept check: *Why is the Barr body more brightly staining in a cell nucleus?*

In addition to this cytological evidence, Lyon was also familiar with examples in which the coat color of a mammal had a variegated pattern. **Figure 4.5b** is a photo of a calico cat, which is a female that is heterozygous for a gene on the X chromosome that can occur as an orange or a black allele. (The white underside is due to a dominant allele in a different gene.) The orange and black patches are randomly distributed in different female individuals. The calico pattern does not occur in male cats, but similar kinds of mosaic patterns have been identified in the female mouse. Lyon suggested that both the Barr body and the calico pattern are the result of X inactivation in the cells of female mammals.

The mechanism of X inactivation, also known as the **Lyon hypothesis,** is schematically illustrated in **Figure 4.6**. This

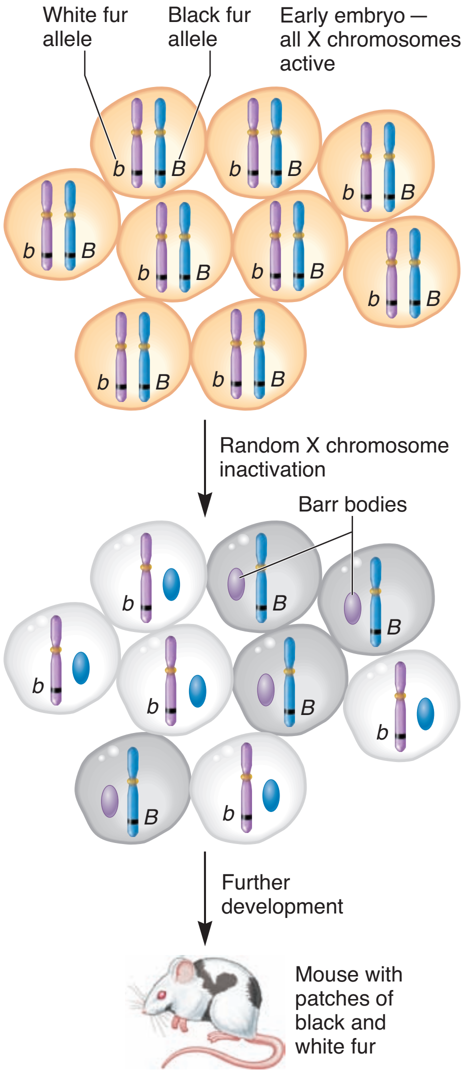

White fur allele
Black fur allele
Early embryo — all X chromosomes active
b
B
Random X chromosome inactivation
Barr bodies
Further development
Mouse with patches of black and white fur

FIGURE 4.6 **The mechanism of X chromosome inactivation.**

Genes→Traits The top of this figure represents a mass of several cells that compose the early embryo. Initially, both X chromosomes are active. At an early stage of embryonic development, random inactivation of one X chromosome occurs in each cell. This inactivation pattern is maintained as the embryo matures into an adult.

Concept check: *At which stage of development does X inactivation initially occur?*

example involves a white and black variegated coat color found in certain strains of mice. As shown here, a female mouse has inherited an X chromosome from its mother that carries an allele conferring white coat color (X^b). The X chromosome from its father carries a black coat color allele (X^B). How can X inactivation explain a variegated coat pattern? Initially, both X chromosomes are active. However, at an early stage of embryonic development, one of the two X chromosomes is randomly inactivated in each somatic cell and becomes a Barr body. For example, one embryonic cell may have the X^B chromosome inactivated. As the embryo continues to grow and mature, this embryonic cell divides and may eventually give rise to billions of cells in the adult animal. The epithelial (skin) cells that are derived from this embryonic cell produce a patch of white fur because the X^B chromosome has been permanently inactivated. Alternatively, another embryonic cell may have the other X chromosome inactivated (i.e., X^b). The epithelial cells derived from this embryonic cell produce a patch of black fur. Because the primary event of X inactivation is a random process that occurs at an early stage of development, the result is an animal with some patches of white fur and other patches of black fur. This is the basis of the variegated phenotype.

During inactivation, the chromosomal DNA becomes highly compacted into a Barr body, so most of the genes on the inactivated X chromosome cannot be expressed. When cell division occurs and the inactivated X chromosome is replicated, both copies remain highly compacted and inactive. Likewise, during subsequent cell divisions, X inactivation is passed along to all future somatic cells.

Mammals Maintain One Active X Chromosome in their Somatic Cells

Since the Lyon hypothesis was confirmed, the genetic control of X inactivation has been investigated further by several laboratories. Research has shown that mammalian cells possess the ability to count their X chromosomes in their somatic cells and allow only one of them to remain active. How was this determined? A key observation came from comparisons of the chromosome composition of people who were born with normal or abnormal numbers of sex chromosomes.

Phenotype	Chromosome Composition	Number of X Chromosomes	Number of Barr Bodies
Normal female	XX	2	1
Normal male	XY	1	0
Turner syndrome (female)	X0	1	0
Triple X syndrome (female)	XXX	3	2
Klinefelter syndrome (male)	XXY	2	1

In normal females, two X chromosomes are counted and one is inactivated, whereas in males, one X chromosome is counted and none inactivated. If the number of X chromosomes exceeds two, as in triple X syndrome, additional X chromosomes are converted to Barr bodies.

X Inactivation in Mammals Depends on the X-Inactivation Center and the *Xist* Gene

Although the genetic control of inactivation is not entirely understood at the molecular level, a short region on the X chromosome called the **X-inactivation center (Xic)** is known to play a critical role. Eeva Therman and Klaus Patau identified the Xic from its key role in X inactivation. The counting of human X chromosomes is accomplished by counting the number of Xics. A Xic must be found on an X chromosome for inactivation to occur. Therman and Patau discovered that if one of the two X chromosomes in a female is missing its Xic due to a chromosome mutation, a cell counts only one Xic and X inactivation does not occur. Having two active X chromosomes is a lethal condition for a human female embryo.

Let's consider how the molecular expression of certain genes controls X inactivation. The expression of a specific gene within the Xic is required for the compaction of the X chromosome into a Barr body. This gene, discovered in 1991, is named *Xist* (for X-inactive specific transcript). The *Xist* gene on the inactivated X chromosome is active, which is unusual because most other genes on the inactivated X chromosome are silenced. The *Xist* gene product is an RNA molecule that does not encode a protein. Instead, the role of the *Xist* RNA is to coat the X chromosome and inactivate it. After coating, other proteins associate with the *Xist* RNA and promote chromosomal compaction into a Barr body.

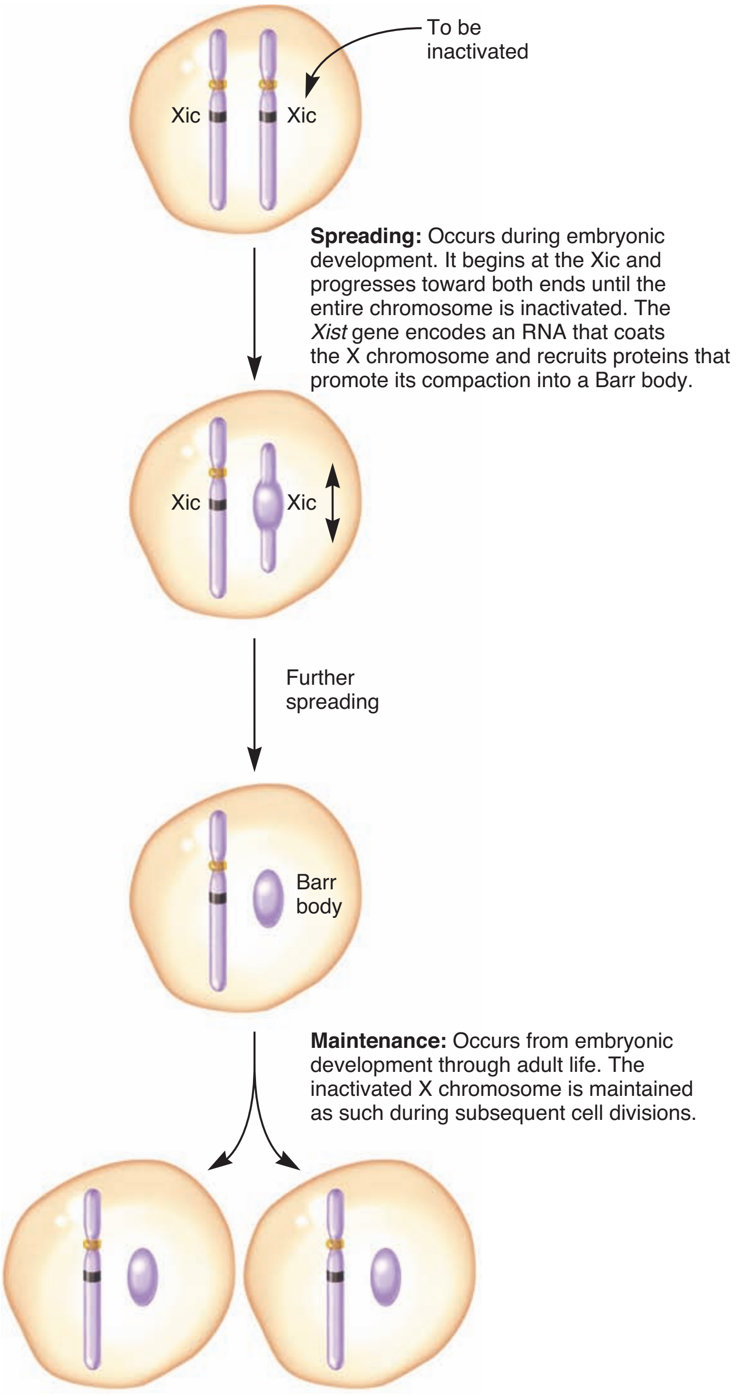
X Inactivation Occurs in Three Phases: Initiation, Spreading, and Maintenance

The process of X inactivation can be divided into three phases: initiation, spreading, and maintenance (**Figure 4.7**). During initiation, which occurs during embryonic development, one of the X chromosomes remains active, and the other is chosen to be inactivated. During the spreading phase, the chosen X chromosome is inactivated. This spreading requires the expression of the *Xist* gene. The *Xist* RNA coats the inactivated X chromosome and recruits proteins that promote compaction. The spreading phase is so named because inactivation begins near the Xic and spreads in both directions along the X chromosome.

Once the initiation and spreading phases occur for a given X chromosome, the inactivated X chromosome is maintained as a Barr body during future cell divisions. When a cell divides, the Barr body is replicated, and both copies remain compacted. This maintenance phase continues from the embryonic stage through adulthood.

Some genes on the inactivated X chromosome are expressed in the somatic cells of adult female mammals. These genes are said to escape the effects of X inactivation. As mentioned, *Xist* is an example of a gene that is expressed from the highly condensed Barr body. In humans, up to a quarter of the genes on the X chromosome may escape inactivation to some degree. Many of these genes occur in clusters. Among these are the pseudoautosomal genes found on both the X and Y chromosomes in the regions of homology, which are described next. Dosage compensation is not necessary for pseudoautosomal genes because they are located on both the X and Y chromosomes.

Initiation: Occurs during embryonic development. The number of X-inactivation centers (Xics) is counted and one of the X chromosomes remains active and the other is targeted for inactivation.



Spreading: Occurs during embryonic development. It begins at the Xic and progresses toward both ends until the entire chromosome is inactivated. The *Xist* gene encodes an RNA that coats the X chromosome and recruits proteins that promote its compaction into a Barr body.

Maintenance: Occurs from embryonic development through adult life. The inactivated X chromosome is maintained as such during subsequent cell divisions.

FIGURE 4.7 **The phases of X inactivation.**

Concept check: Which of these phases occurs in an adult female?

4.2 REVIEWING THE KEY CONCEPTS

- Dosage compensation often occurs in species that differ in their sex chromosomes (see Table 4.1).
- In mammals, the process of X inactivation in females compensates for the single X chromosome found in males. The inactivated X chromosome is called a Barr body. The process can lead to a variegated phenotype, such as a calico cat (see Figure 4.5).
- After it occurs during embryonic development, the pattern of X inactivation is maintained when cells divide (see Figure 4.6).
- X inactivation is controlled by the X-inactivation center (Xic) that contains the Xist gene. X inactivation occurs as initiation, spreading, and maintenance phases (see Figure 4.7).

4.2 COMPREHENSION QUESTIONS

1. In fruit flies, dosage compensation is achieved by
   a. X inactivation.
   b. turning up the expression of genes on the single X chromosome in the male twofold.
   c. turning down the expression of genes on the two X chromosomes in the female to one half.
   d. all of the above.
2. According to the Lyon hypothesis,
   a. one of the X chromosomes is converted to a Barr body in somatic cells of female mammals.
   b. one of the X chromosomes is converted to a Barr body in all cells of female mammals.
   c. both of the X chromosomes are converted to a Barr body in somatic cells of female mammals.
   d. both of the X chromosomes are converted to a Barr body in all cells of female mammals.
3. Which of the following is *not* a phase of X inactivation?
   a. Initiation
   b. Spreading
   c. Maintenance
   d. Erasure

4.3 PROPERTIES OF THE X AND Y CHROMOSOME IN MAMMALS

As we discussed at the beginning of this chapter, sex determination in mammals is determined by the presence of the Y chromosome, which carries the *Sry* gene. The X and Y chromosomes also differ in other ways. The X chromosome is typically much larger than the Y and carries more genes. For example, in humans, researchers estimate that the X chromosome carries about 1200 to 1500 genes, whereas the Y chromosome has 80 to 200 genes. Genes that are found on only one sex chromosome but not both are called **sex-linked genes.** Those on the X chromosome are termed **X-linked genes** and those on the Y chromosome are termed **Y-linked genes,** or **holandric genes.**

Besides sex-linked genes, the X and Y chromosomes also contain short regions of homology where the X and Y chromosomes carry the same genes, which are called **pseudoautosomal genes.** In addition to several smaller regions, the human

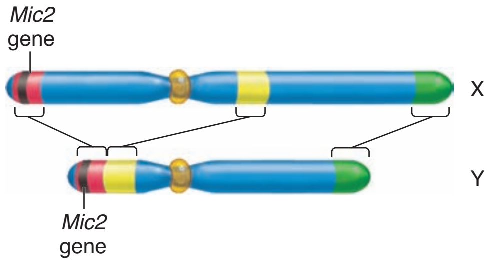


FIGURE 4.8 **A comparison of the homologous and nonhomologous regions of the X and Y chromosome in humans.** The brackets show three regions of homology between the X and Y chromosome. A few pseudoautosomal genes, such as *Mic2*, are found on both the X and Y chromosomes in these small regions of homology.

Concept check: *Why are the homologous regions of the X and Y chromosome important during meiosis?*

sex chromosomes have three homologous regions (**Figure 4.8**). These regions, which are evolutionarily related, promote the necessary pairing of the X and Y chromosomes that occurs during meiosis I of spermatogenesis. Relatively few genes are located in these homologous regions. One example is a human gene called *Mic2*, which encodes a cell surface antigen. The *Mic2* gene is found on both the X and Y chromosomes. It follows a pattern of inheritance called **pseudoautosomal inheritance.** The term *pseudoautosomal* refers to the idea that the inheritance pattern of the *Mic2* gene is the same as the inheritance pattern of a gene located on an autosome even though the *Mic2* gene is actually located on the sex chromosomes. As in autosomal inheritance, males have two copies of pseudoautosomally inherited genes, and they can transmit the genes to both daughters and sons.

By comparison, genes that are found only on the X or Y chromosome exhibit transmission patterns that are quite different from genes located on an autosome. A Y-linked inheritance pattern is very distinctive—the gene is transmitted only from fathers to sons. By comparison, transmission patterns involving X-linked genes are more complex because females inherit two X chromosomes whereas males receive only one X chromosome from their mother. We will consider the complexities of X-linked inheritance patterns next.

4.3 REVIEWING THE KEY CONCEPTS

- Chromosomes that differ between males and females are termed sex chromosomes and carry sex-linked genes.
- X-linked genes are found only on the X chromosome, whereas Y-linked genes are found only on the Y chromosome. Pseudoautosomal genes are found on both the X and Y chromosomes in regions of homology (see Figure 4.8).

4.3 COMPREHENSION QUESTIONS

1 A Y-linked gene is passed from
  a. father to son.
  b. father to daughter.
  c. father to daughter or son.
  d. mother to son.

4.4 TRANSMISSION PATTERNS FOR X-LINKED GENES

At the beginning of this chapter, we discussed how sex determination in certain species is controlled by sex chromosomes. In fruit flies and mammals, a female is XX, whereas a male is XY. The inheritance pattern of X-linked genes, known as **X-linked inheritance,** shows certain distinctive features. For example, males transmit X-linked genes only to their daughters, and sons receive their X-linked genes from their mothers. The term **hemizygous** is used to describe the single copy of an X-linked gene in the male. A male mammal or fruit fly is said to be hemizygous for X-linked genes. Because males of certain species, such as humans, have a single copy of the X chromosome, another distinctive feature of X-linked inheritance is that males are more likely to be affected by rare, recessive X-linked disorders. We will consider the medical implications of X-linked inheritance in Chapter 23. In this section, we will examine X-linked inheritance in fruit flies and mammals.

Morgan's Experiments Showed a Connection Between a Genetic Trait and the Inheritance of a Sex Chromosome in *Drosophila*

In the early 1900s, Thomas Hunt Morgan carried out the first study that confirmed the location of a gene on a particular chromosome. In this experiment, he showed that a gene affecting eye color in fruit flies is located on the X chromosome. Morgan was trained as an embryologist, and much of his early research involved descriptive and experimental work in that field. He was particularly interested in ways that organisms change. He wrote, "The most distinctive problem of zoological work is the change in form that animals undergo, both in the course of their development from the egg (embryology) and in their development in time (evolution)." Throughout his life, he usually had dozens of different experiments going on simultaneously, many of them unrelated to each other. He jokingly said there are three kinds of experiments—those that are foolish, those that are damn foolish, and those that are worse than that!

In one of his most famous studies, Morgan engaged one of his graduate students to rear the fruit fly *Drosophila melanogaster* in the dark, hoping to produce flies whose eyes would atrophy from disuse and disappear in future generations. Even after many consecutive generations, however, the flies appeared to have no noticeable changes despite repeated attempts at inducing mutations by treatments with agents such as X-rays and radium. After 2 years, Morgan finally obtained an interesting result when a true-breeding line of *Drosophila* produced a male fruit fly with white eyes rather than the normal red eyes. Because this had been a true-breeding line of flies, this white-eyed male must have arisen from a new mutation that converted a red-eye allele (denoted w^+) into a white-eye allele (denoted w). Morgan is said to have carried this fly home with him in a jar, put it by his bedside at night while he slept, and then taken it back to the laboratory during the day.

Much like Mendel, Morgan studied the inheritance of this white-eye trait by making crosses and quantitatively analyzing their outcome. In the experiment described in **Figure 4.9**, he began with his white-eyed male and crossed it to a true-breeding red-eyed female. All of the F_1 offspring had red eyes, indicating that red is dominant to white. The F_1 offspring were then mated to each other to obtain an F_2 generation.

▶ THE GOAL

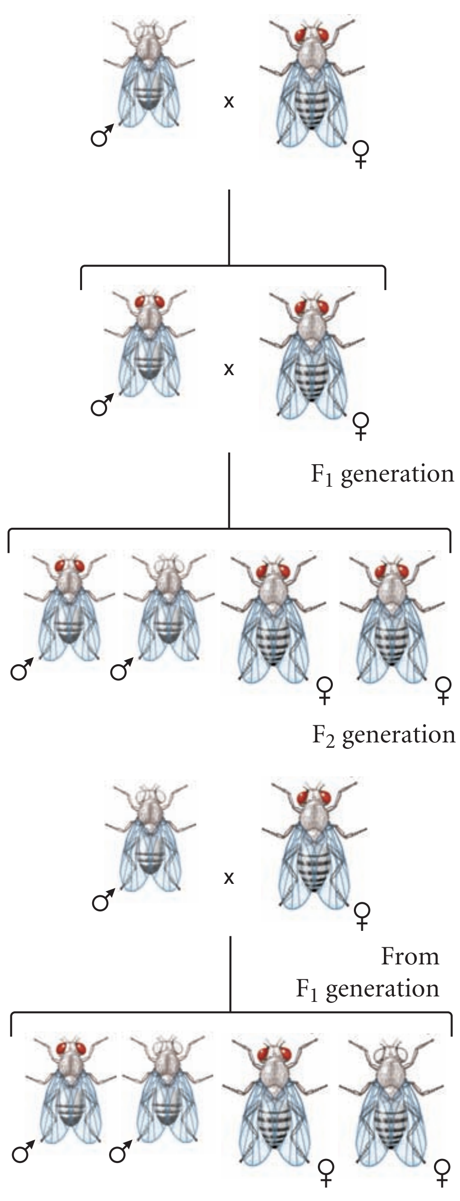
This is an example of discovery-based science rather than hypothesis testing. In this case, a quantitative analysis of genetic crosses may reveal the inheritance pattern for the white-eye allele.

▶ ACHIEVING THE GOAL — FIGURE 4.9 Inheritance pattern of an X-linked trait in fruit flies.

Concept check: *What is the key result that suggests an X-linked inheritance pattern?*

Starting material: A true-breeding line of red-eyed fruit flies plus one white-eyed male fly that was discovered in Morgan's collection of flies.

Experimental level | **Conceptual level**



1. Cross the white-eyed male to a true-breeding red-eyed female.

$X^wY \times X^{w^+}X^{w^+}$

2. Record the results of the F_1 generation. This involves noting the eye color and sexes of many offspring.

$X^{w^+}Y$ male offspring and $X^{w^+}X^w$ female offspring, both with red eyes

$X^{w^+}Y \times X^{w^+}X^w$

3. Cross F_1 offspring with each other to obtain F_2 offspring. Also record the eye color and sex of the F_2 offspring.

$1\ X^{w^+}Y : 1\ X^wY : 1\ X^{w^+}X^{w^+} : 1\ X^{w^+}X^w$
1 red-eyed male : 1 white-eyed male :
2 red-eyed females

4. In a separate experiment, perform a testcross between a white-eyed male and a red-eyed female from the F_1 generation. Record the results.

$X^wY \times X^{w^+}X^w$

$1\ X^{w^+}Y : 1\ X^wY : 1\ X^{w^+}X^w : 1\ X^wX^w$
1 red-eyed male : 1 white-eyed male :
1 red-eyed female : 1 white-eyed female

▶ THE DATA

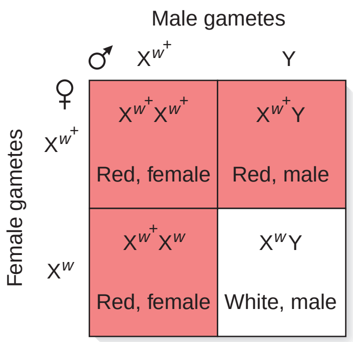
Cross	*Results*	
Original white-eyed male to red-eyed females	F_1 generation:	All red-eyed flies
F_1 males to F_1 females	F_2 generation:	2459 red-eyed females 1011 red-eyed males 0 white-eyed females 782 white-eyed males
White-eyed males to F_1 females	Testcross:	129 red-eyed females 132 red-eyed males 88 white-eyed females 86 white-eyed males

Data from T. H. Morgan (1910) Sex limited inheritance in *Drosophila. Science* 32, 120–122.

▶ INTERPRETING THE DATA

As seen in the data table, the F_2 generation consisted of 2459 red-eyed females, 1011 red-eyed males, and 782 white-eyed males. Most notably, no white-eyed female offspring were observed in the F_2 generation. These results suggested that the pattern of transmission from parent to offspring depends on the sex of the offspring and on the alleles that they carry. As shown in the Punnett square here, the data are consistent with the idea that the eye color alleles are located on the X chromosome:

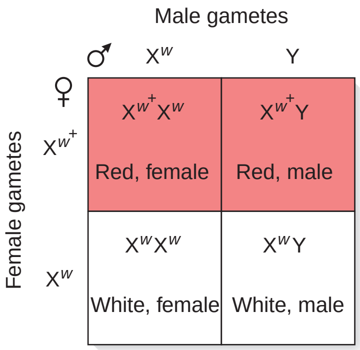
F_1 male is $X^{w^+}Y$
F_1 female is $X^{w^+}X^{w}$



The Punnett square predicts that the F_2 generation will not have any white-eyed females. This prediction was confirmed experimentally. These results indicated that the eye color alleles are located on the X chromosome. As mentioned earlier, genes that are physically located within the X chromosome are called **X-linked genes,** or **X-linked alleles.** However, it should also be pointed out that the experimental ratio in the F_2 generation of red eyes to white eyes is (2459 + 1011):782, which equals 4.4:1. This ratio deviates significantly from the predicted ratio of 3:1. How can this discrepancy be explained? Later work revealed that the lower-than-expected number of white-eyed flies is due to their decreased survival rate.

Morgan also conducted a **testcross** (see step 4, Figure 4.9) in which an individual with a dominant phenotype and unknown genotype is crossed to an individual with a recessive phenotype. In this case, he mated an F_1 red-eyed female to a white-eyed male. This cross produced red-eyed males and females, and white-eyed males and females, in approximately equal numbers. The testcross data are also consistent with an X-linked pattern of inheritance. As shown in the following Punnett square, the testcross predicts a 1:1:1:1 ratio:

Testcross:
Male is $X^{w}Y$
F_1 female is $X^{w^+}X^{w}$



The observed data were 129:132:88:86, which is a ratio of 1.5:1.5:1:1. Again, the lower-than-expected numbers of white-eyed males and females can be explained by a lower survival rate for white-eyed flies. In his own interpretation, Morgan concluded that red eye color and X (a sex factor that is present in two copies in the female) are combined and have never existed apart. In other words, this gene for eye color is on the X chromosome. In 1933, Morgan received the Nobel Prize in physiology or medicine.

Calvin Bridges, a graduate student in the laboratory of Morgan, also examined the transmission of X-linked traits. In his crosses, he occasionally obtained offspring that had unexpected phenotypes and abnormalities in sex chromosome composition. For example, in a cross between a white-eyed female and a red-eyed male, he occasionally observed a male offspring with red eyes. This event can be explained by nondisjunction, which is described in Chapter 8 (see Figure 8.21). In this example, the rare male offspring with red eyes was produced by a sperm carrying the X-linked red allele and by an egg that underwent nondisjunction and did not receive an X chromosome. The resulting offspring would be a male without a Y chromosome. (As shown earlier in Figure 4.1, the number of X chromosomes determines sex in fruit flies.) Together, the work of Morgan and Bridges provided an impressive body of evidence confirming the idea that traits following an X-linked pattern of inheritance are governed by genes that are physically located on the X chromosome. Bridges wrote, “There can be no doubt that the complete parallelism between the unique behavior of chromosomes and the behavior of sex-linked genes and sex in this case means that the sex-linked genes are located in and borne by the X chromosomes.” An example of Bridges’ work is described in solved problem S5 at the end of this chapter.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

The Inheritance Pattern of X-Linked Genes Can Be Revealed by Reciprocal Crosses

X-linked patterns are also observed in mammalian species. As an example, let's consider a human disease known as Duchenne muscular dystrophy (DMD), which was first described by the French neurologist Guillaume Duchenne in the 1860s. Affected individuals show signs of muscle weakness as early as age 3. The disease gradually weakens the skeletal muscles and eventually affects the heart and breathing muscles. Survival is rare beyond the early 30s. The gene for DMD, found on the X chromosome, encodes a protein called dystrophin that is required inside muscle cells for structural support. Dystrophin is thought to strengthen muscle cells by anchoring elements of the internal cytoskeleton to the plasma membrane. Without it, the plasma membrane becomes permeable and may rupture.

DMD is inherited in an **X-linked recessive pattern**—the allele causing the disease is recessive and located on the X chromosome. In the pedigree shown in **Figure 4.10**, several males are affected by this disorder, as indicated by filled squares. The mothers of these males are presumed heterozygotes for this X-linked recessive allele. This recessive disorder is very rare among females because daughters would have to inherit a copy of the mutant allele from their mother and a copy from an affected father.

X-linked muscular dystrophy has also been found in certain breeds of dogs such as golden retrievers (**Figure 4.11a**). Like humans, the mutation occurs in the dystrophin gene, and the symptoms include severe weakness and muscle atrophy that begin at about 6 to 8 weeks of age. Many dogs that inherit this disorder die within the first year of life, though some can live 3 to 5 years and reproduce.

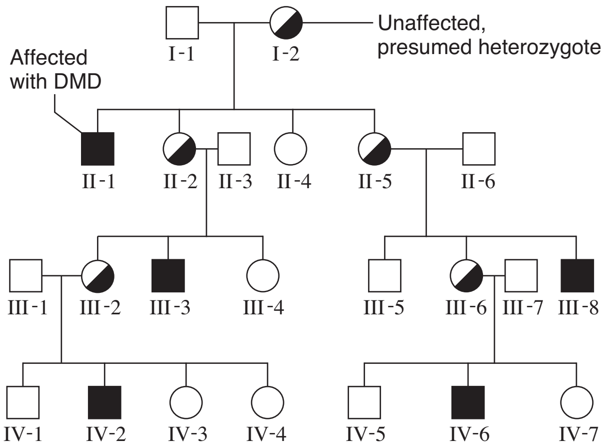
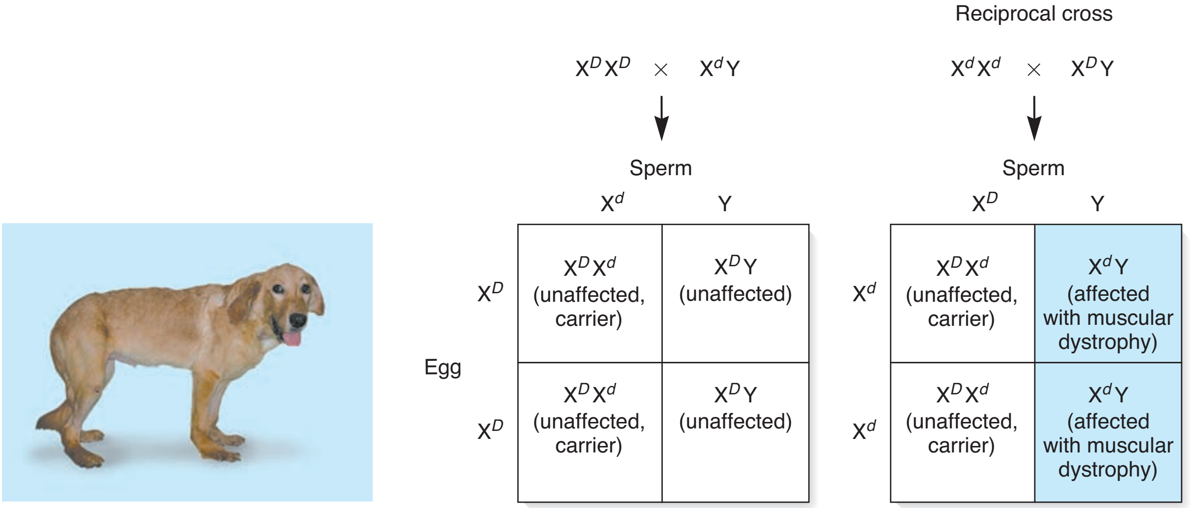


FIGURE 4.10 **A human pedigree for Duchenne muscular dystrophy, an X-linked recessive trait.** Affected individuals are shown with filled symbols. Females who are unaffected with the disease but have affected sons are presumed to be heterozygous carriers, as shown with half-filled symbols.

Concept check: What features of this pedigree indicate that the allele for Duchenne muscular dystrophy is X-linked?

Figure 4.11b (left side) considers a cross between an unaffected female dog with two copies of the wild-type gene and a male dog with muscular dystrophy that carries the mutant allele and has survived to reproductive age. When setting up a Punnett square involving X-linked traits, we must consider the alleles on the X chromosome as well as the observation that males may transmit a Y chromosome instead of the X chromosome. The male



(a) Male golden retriever with X-linked muscular dystrophy

(b) Examples of X-linked muscular dystrophy inheritance patterns

FIGURE 4.11 **X-linked muscular dystrophy in dogs.** **(a)** The male golden retriever shown here has the disease. **(b)** The left side shows a cross between an unaffected female and an affected male. The right shows a reciprocal cross between an affected female and an unaffected male. *D* represents the normal allele for the dystrophin gene, and *d* is the mutant allele that causes a defect in dystrophin function.

Concept check: Explain why the reciprocal cross yields a different result compared to the first cross.

makes two types of gametes: one that carries the X chromosome and one that carries the Y. The Punnett square must also include the Y chromosome even though this chromosome does not carry any X-linked genes. The X chromosomes from the female and male are designated with their corresponding alleles. When the Punnett square is filled in, it predicts the X-linked genotypes and sexes of the offspring. As seen on the left side of Figure 4.11b, none of the offspring from this cross are affected with the disorder, although all female offspring are carriers.

The right side of Figure 4.11b shows a **reciprocal cross**—a second cross in which the sexes and phenotypes are reversed. In this case, an affected female animal is crossed to an unaffected male. This cross produces female offspring that are carriers and all male offspring will be affected with muscular dystrophy.

When comparing the two Punnett squares, the outcome of the reciprocal cross yielded different results. This is expected of X-linked genes, because the male transmits the gene only to female offspring, whereas the female transmits an X chromosome to both male and female offspring. Because the male parent does not transmit the X chromosome to his sons, he does not contribute to their X-linked phenotypes. This explains why X-linked traits do not behave equally in reciprocal crosses. Experimentally, the observation that reciprocal crosses do not yield the same results is an important clue that a trait may be X-linked.

4.4 REVIEWING THE KEY CONCEPTS

- Morgan's work showed that a gene affecting eye color in fruit flies is inherited on the X chromosome (see Figure 4.9).
- X-linked inheritance is also observed in mammals. An example is Duchenne muscular dystrophy, which is an X-linked recessive disorder. For X-linked alleles, reciprocal crosses yield different results (see Figures 4.10, 4.11).

4.4 COMPREHENSION QUESTIONS

1. In Morgan's experiments, a key observation of the F_2 generation that suggested an X-linked pattern of inheritance was
   a. a 3:1 ratio of red-eyed to white-eyed flies.
   b. the only white-eyed flies were males.
   c. the only white-eyed flies were females.
   d. the only red-eyed flies were males.
2. Hemophilia is a blood clotting disorder in humans that follows an X-linked recessive pattern of inheritance. A man with hemophilia and woman without hemophilia have a daughter with hemophilia. If you let *H* represent the normal allele and *h* the hemophilia allele, what are the genotypes of the parents?
   a. Mother is X^HX^h and father is $X^h Y$.
   b. Mother is X^hX^h and father is $X^h Y$.
   c. Mother is X^hX^h and father is $X^H Y$.
   d. Mother is X^HX^h and father is $X^H Y$.
3. A cross is made between a white-eyed female fruit fly and a red-eyed male. What would be the reciprocal cross?
   a. Female is X^wX^w and male is $X^w Y$.
   b. Female is $X^{w+}X^{w+}$ and father is $X^{w+} Y$.
   c. Mother is $X^{w+}X^{w+}$ and father is $X^w Y$.
   d. Mother is X^wX^w and father is $X^{w+} Y$.

KEY TERMS

Page 71. sex determination
Page 72. sex chromosomes, heterogametic sex, homogametic sex, autosomes, haplodiploid system
Page 73. protandrous hermaphrodites
Page 74. dioecious
Page 75. dosage compensation, X inactivation
Page 76. Barr body, Lyon hypothesis
Page 77. X-inactivation center (Xic)
Page 78. sex-linked genes, X-linked genes, Y-linked genes, holandric genes, pseudoautosomal genes
Page 79. pseudoautosomal inheritance, X-linked inheritance, hemizygous
Page 81. X-linked genes, X-linked alleles, testcross
Page 82. X-linked recessive pattern
Page 83. reciprocal cross

CHAPTER SUMMARY

4.1 Mechanisms of Sex Determination Among Various Species

- In the X-Y, X-0, and Z-W systems, sex is determined by the presence and number of particular sex chromosomes (see Figure 4.1).
- In the haplodiploid system, sex is determined by the number of sets of chromosomes (see Figure 4.2).
- In some species, such as alligators and clownfish, sex is determined by environmental factors (see Figure 4.3).
- Dioecious plants exist as individuals that produce only pollen and those that produce only eggs. Sex chromosomes sometimes determine sex in these species (see Figure 4.4).

4.2 Dosage Compensation and X Inactivation in Mammals

- Dosage compensation often occurs in species that differ in their sex chromosomes (see Table 4.1).
- In mammals, the process of X inactivation in females compensates for the single X chromosome found in males. The

inactivated X chromosome is called a Barr body. The process can lead to a variegated phenotype, such as a calico cat (see Figure 4.5).

- After it occurs during embryonic development, the pattern of X inactivation is maintained when cells divide (see Figures 4.6).
- X inactivation is controlled by the X-inactivation center (Xic) that contains the *Xist* gene. X inactivation occurs as initiation, spreading, and maintenance phases (see Figure 4.7).

4.3 Properties of the X and Y Chromosome in Mammals

- Chromosomes that differ between males and females are termed sex chromosomes and carry sex-linked genes.
- X-linked genes are found only on the X chromosome, whereas Y-linked genes are found only on the Y chromosome. Pseudoautosomal genes are found on both the X and Y chromosomes in regions of homology (see Figure 4.8).

4.4 Transmission Patterns for X-Linked Genes

- Morgan's work showed that a gene affecting eye color in fruit flies is inherited on the X chromosome (see Figure 4.9).
- X-linked inheritance is also observed in mammals. An example is Duchenne muscular dystrophy, which is an X-linked recessive disorder. For X-linked alleles, reciprocal crosses yield different results (see Figures 4.10, 4.11).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. A human male has an X chromosome that is missing its Xic. Is this caused by a new mutation (one that occurred during gametogenesis), or could this mutation have occurred in an earlier generation and be found in the somatic cells of one of his parents? Explain your answer. How would this mutation affect his ability to produce viable offspring?

Answer: The missing Xic must be due to a new mutation that occurred during oogenesis. A male inherits his X chromosome from his mother. If his mother had inherited an X chromosome with a missing Xic, both of her X chromosomes would remain active, and this is a lethal condition in humans. For this same reason, the male with the missing Xic could not produce living daughters, but he could produce normal sons.

S2. In humans, why are X-linked recessive traits more likely to occur in males than females?

Answer: Because a male is hemizygous for X-linked traits, the phenotypic expression of X-linked traits depends on only a single copy of the gene. When a male inherits a recessive X-linked allele, he automatically exhibits the trait because he does not have another copy of the gene on the corresponding Y chromosome. This phenomenon is particularly relevant to the inheritance of recessive X-linked alleles that cause human disease. (Some examples will be described in Chapter 23.)

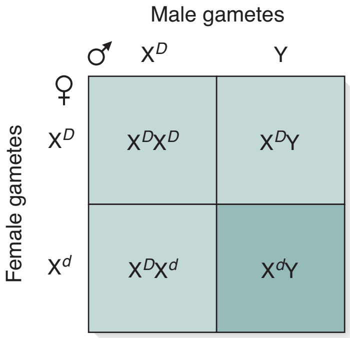
S3. An unaffected woman (i.e., without disease symptoms) who is heterozygous for the X-linked allele causing Duchenne muscular dystrophy has children with a man with a normal allele. What are the probabilities of the following combinations of offspring?

A. An unaffected son

B. An unaffected son or daughter

C. A family of three children, all of whom are affected

Answer: The first thing we must do is construct a Punnett square to determine the outcome of the cross. *D* represents the normal allele, and *d* is the recessive allele causing Duchenne muscular dystrophy. The mother is heterozygous, and the father has the normal allele.



♂ / ♀	X^D	Y
X^D	X^DX^D	X^DY
X^d	X^DX^d	X^dY

Phenotype ratio is

2 normal daughters :
1 normal son :
1 affected son

A. There are four possible children, one of whom is an unaffected son. Therefore, the probability of an unaffected son is 1/4.

B. Use the sum rule: 1/4 + 1/2 = 3/4.

C. You could use the product rule because there would be three offspring in a row with the disorder: (1/4)(1/4)(1/4) = 1/64 = 0.016 = 1.6%.

S4. Red-green color blindness is inherited as a recessive X-linked trait. What are the following probabilities?

A. A woman with phenotypically normal parents and a color-blind brother will have a color-blind son. Assume that she has no previous children.

B. The next child of a phenotypically normal woman, who has already had one color-blind son, will be a color-blind son.

C. The next child of a phenotypically normal woman, who has already had one color-blind son, and who is married to a color-blind man, will have a color-blind daughter.

Answer:

A. The woman's mother must have been a heterozygote. So there is a 50% chance that the woman is a carrier. If she has children, 1/4 (i.e., 25%) will be affected sons if she is a carrier. However, there is only a 50% chance that she is a carrier. We multiply 50% times 25%, which equals 0.5 × 0.25 = 0.125, or a 12.5% chance.

B. If she already had a color-blind son, then we know she must be a carrier, so the chance is 25%.

C. The woman is heterozygous and her husband is hemizygous for the color-blind allele. This couple will produce 1/4 offspring that are color-blind daughters. The rest are 1/4 carrier daughters, 1/4 normal sons, and 1/4 color-blind sons. Answer is 25%.

S5. To test the chromosome theory of inheritance, which is described in Chapter 3, Calvin Bridges made crosses involving the inheritance of X-linked traits. One of his experiments concerned two different X-linked genes affecting eye color and wing size. For the eye color gene, the red-eye allele (w^+) is dominant to the white-eye allele (w). A second X-linked trait is wing size; the allele called miniature is recessive to the normal allele. In this case, m represents the miniature allele and m^+ the normal allele. A male fly carrying a miniature allele on its single X chromosome has small (miniature) wings. A female must be homozygous, mm, in order to have miniature wings.

Bridges made a cross between $X^{w,m^+}X^{w,m^+}$ female flies (white eyes and normal wings) to $X^{w^+,m}Y$ male flies (red eyes and miniature wings). He then examined the eyes, wings, and sexes of thousands of offspring. Most of the offspring were females with red eyes and normal wings, and males with white eyes and normal wings. On rare occasions (approximately 1 out of 1700 flies), however, he also obtained female offspring with white eyes or males with red eyes. He also noted the wing shape in these flies and then cytologically examined their chromosome composition using a microscope. The following results were obtained:

Offspring	Eye Color	Wing Size	Sex Chromosomes
Expected females	Red	Normal	XX
Expected males	White	Normal	XY
Unexpected females (rare)	White	Normal	XXY
Unexpected males (rare)	Red	Miniature	X0

Data from: Bridges, C. B. (1916) "Non-disjunction as proof of the chromosome theory of heredity," *Genetics 1*, 1–52, 107–163.

Explain these data.

Answer: Remember that in fruit flies, the number of X chromosomes (not the presence of the Y chromosome) determines sex. As seen in the data, the flies with unexpected phenotypes were abnormal in their sex chromosome composition. The white-eyed female flies were due to the union between an abnormal XX female gamete and a normal Y male gamete. Likewise, the unexpected male offspring contained only one X chromosome and no Y. These male offspring were due to the union between an abnormal egg without any X chromosome and a normal sperm containing one X chromosome. The wing size of the unexpected males was a particularly significant result. The red-eyed males showed a miniature wing size. As noted by Bridges, this means they inherited their X chromosome from their father rather than their mother. This observation provided compelling evidence that the inheritance of the X chromosome correlates with the inheritance of particular traits.

Conceptual Questions

C1. In the X-Y and Z-W systems of sex determination, which sex is the heterogametic sex, the male or female?

C2. Let's suppose that a gene affecting pigmentation is found on the X chromosome (in mammals or insects) or the Z chromosome (in birds) but not on the Y or W chromosome. It is found on an autosome in bees. This gene exists in two alleles, D (dark), which is dominant to d (light). What would be the phenotypic results of crosses between a true-breeding dark female and true-breeding light male, and the reciprocal crosses involving a true-breeding light female and true-breeding dark male, in the following species? Refer back to Figures 4.1 and 4.2 for the mechanism of sex determination in these species.

A. Birds

B. *Drosophila*

C. Bees

D. Humans

C3. Assuming that such a fly would be viable, what would be the sex of a fruit fly with the following chromosomal composition?

A. One X chromosome and two sets of autosomes

B. Two X chromosomes, one Y chromosome, and two sets of autosomes

C. Two X chromosomes and four sets of autosomes

D. Four X chromosomes, two Y chromosomes, and four sets of autosomes

C4. What would be the sex of a human with the following numbers of sex chromosomes?

A. XXX

B. X (also described as X0)

C. XYY

D. XXY

C5. With regard to the numbers of sex chromosomes, explain why dosage compensation is necessary.

C6. What is a Barr body? How is its structure different from that of other chromosomes in the cell? How does the structure of a Barr body affect the level of X-linked gene expression?

C7. Among different species, describe three distinct strategies for accomplishing dosage compensation.

C8. Describe when X inactivation occurs and how this leads to phenotypic results at the organism level. In your answer, you should explain why X inactivation causes results such as variegated coat patterns in mammals. Why do two different calico cats have their patches of orange and black fur in different places? Explain whether or not a variegated coat pattern due to X inactivation could occur in marsupials.

C9. Describe the molecular process of X inactivation. This description should include the three phases of inactivation and the role of the Xic. Explain what happens to X chromosomes during embryogenesis, in adult somatic cells, and during oogenesis.

C10. On rare occasions, an abnormal human male is born who is somewhat feminized compared with normal males. Microscopic examination of the cells of one such individual revealed that he has a single Barr body in each cell. What is the chromosomal composition of this individual?

C11. How many Barr bodies would you expect to find in humans with the following abnormal compositions of sex chromosomes?

A. XXY

B. XYY

C. XXX

D. X0 (a person with just a single X chromosome)

C12. Certain forms of human color blindness are inherited as an X-linked recessive trait. Hemizygous males are color-blind, but heterozygous females are not. However, heterozygous females sometimes have partial color blindness.

A. Discuss why heterozygous females may have partial color blindness.

B. Doctors identified an unusual case in which a heterozygous female was color-blind in her right eye but had normal color vision in her left eye. Explain how this might have occurred.

C13. A black female cat (X^BX^B) and an orange male cat (X^OY) were mated to each other and produced a male cat that was calico. Which sex chromosomes did this male offspring inherit from its mother and father? Remember that the presence of the Y chromosome determines maleness in mammals.

C14. What is the spreading phase of X inactivation? Why do you think it is called a spreading phase? Discuss the role of the *Xist* gene in the spreading phase of X inactivation.

C15. A human disease known as vitamin D-resistant rickets is inherited as an X-linked dominant trait. If a male with the disease produces children with a female who does not have the disease, what is the expected ratio of affected and unaffected offspring?

C16. Hemophilia is a X-linked recessive trait in humans. If a heterozygous woman has children with an unaffected man, what are the odds of the following combinations of children?

A. An affected son

B. Four unaffected offspring in a row

C. An unaffected daughter or son

D. Two out of five offspring that are affected

C17. Incontinentia pigmenti is a rare, X-linked dominant disorder in humans characterized by swirls of pigment in the skin. If an affected female, who had an unaffected father, has children with an unaffected male, what would be the predicted ratios of affected and unaffected sons and daughters?

Application and Experimental Questions

E1. Discuss the types of experimental observations that Mary Lyon brought together in proposing her hypothesis concerning X inactivation. In your own words, explain how these observations were consistent with her hypothesis.

E2. The *Mic2* gene in humans is present on both the X and Y chromosomes. Let's suppose the *Mic2* gene exists in a dominant *Mic2* allele, which results in normal surface antigen, and a recessive *mic2* allele, which results in defective surface antigen production. Using molecular techniques, it is possible to identify homozygous and heterozygous individuals. By following the transmission of the *Mic2* and *mic2* alleles in a large human pedigree, would it be possible to distinguish between pseudoautosomal inheritance and autosomal inheritance? Explain your answer.

E3. In Morgan's experiments, which result do you think is the most convincing piece of evidence pointing to X-linkage of the eye color gene? Explain your answer.

E4. In his original studies of Figure 4.9, Morgan first suggested that the original white-eyed male had two copies of the white-eye allele. In this problem, let's assume that he meant the fly was X^wY^w instead of X^wY. Are his data in Figure 4.9 consistent with this hypothesis? What crosses would need to be made to rule out the possibility that the Y chromosome carries a copy of the eye color gene?

E5. How would you set up crosses to determine if a gene was Y-linked versus X-linked?

E6. Occasionally during meiosis, a mistake can happen whereby a gamete may receive zero or two sex chromosomes rather than one. Calvin Bridges made a cross between white-eyed female flies and red-eyed male flies. As you would expect, most of the offspring were red-eyed females and white-eyed males. On rare occasions, however, he found a white-eyed female or a red-eyed male. These rare flies were not due to new gene mutations but instead were due to mistakes during meiosis in the parent flies. Consider the mechanism of sex determination in fruit flies and propose how this could happen. In your answer, describe the sex chromosome composition of these rare flies.

E7. White-eyed flies have a lower survival rate than red-eyed flies. Based on the data in Figure 4.9, what percentage of white-eyed flies survived compared with red-eyed flies, assuming 100% survival of red-eyed flies?

E8. A cross was made between female flies with white eyes and miniature wings (both X-linked recessive traits) to male flies with red eyes and normal wings. On rare occasions, female offspring were produced with white eyes. If we assume these females are due to errors in meiosis, what would be the most likely chromosomal composition of such flies? What would be their wing shape?

E9. Experimentally, how do you think researchers were able to determine that the Y chromosome causes maleness in mammals, whereas the ratio of X chromosomes to the sets of autosomes causes sex determination in fruit flies?

Questions for Student Discussion/Collaboration

1. Sex determination among different species seems to be caused by a wide variety of mechanisms. With regard to evolution, discuss why you think this has happened. During the evolution of life on Earth over the past 4 billion years, do you think the phenomenon of opposite sexes is a relatively recent event?

2. In Figure 4.9, Morgan obtained a white-eyed male fly in a population containing many red-eyed flies that he thought were true-breeding. As mentioned in the experiment, he crossed this fly with several red-eyed sisters, and all the offspring had red eyes. But actually this is not quite true. Morgan observed 1237 red-eyed flies and 3 white-eyed males. Provide two or more explanations why he obtained 3 white-eyed males in the F_1 generation.

Answers to Comprehension Questions

4.1: d, a, b

4.2: b, a, d

4.3: a

4.4: b, a, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

5

CHAPTER OUTLINE

5.1 Overview of Simple Inheritance Patterns
5.2 Dominant and Recessive Alleles
5.3 Environmental Effects on Gene Expression
5.4 Incomplete Dominance, Overdominance, and Codominance
5.5 Sex-Influenced and Sex-Limited Inheritance
5.6 Lethal Alleles
5.7 Pleiotropy
5.8 Gene Interactions

Inheritance patterns and alleles. *In the petunia, multiple alleles can result in flowers with several different colors, such as the three shown here.*

EXTENSIONS OF MENDELIAN INHERITANCE

The term **Mendelian inheritance** describes inheritance patterns that obey two laws: the law of segregation and the law of independent assortment. Until now, we have mainly considered traits that are affected by a single gene that is found in two different alleles. In these cases, one allele is dominant to the other. This type of inheritance is sometimes called **simple Mendelian inheritance** because the observed ratios in the offspring readily obey Mendel's laws. For example, when two different true-breeding pea plants are crossed (e.g., tall and dwarf), and the F_1 generation is allowed to self-fertilize, the F_2 generation shows a 3:1 phenotypic ratio of tall to dwarf offspring. In Chapter 4, we also considered patterns of inheritance when genes are found on sex chromosomes.

In this chapter, we will extend our understanding of Mendelian inheritance in several ways. First, we will take a closer look at simple dominant/recessive patterns of inheritance in order to understand why some alleles are recessive whereas others are dominant. We will then consider situations in which alleles do not show a simple dominant/recessive relationship. Geneticists have discovered an amazing diversity of mechanisms by which alleles affect the outcome of traits. Many alleles don't produce the ratios of offspring that are expected from a simple Mendelian relationship. This does not mean that Mendel was wrong. Rather, the inheritance patterns of many traits are more complex and interesting than he had realized. In this chapter, we will examine how the outcome of a trait may be influenced by a variety of factors, such as the level of protein expression, the sex of the individual, the presence of multiple alleles of a given gene, and environmental effects. At the end of this chapter, we will explore how two different genes can contribute to the outcome of a single trait. Later, in Chapters 6 and 7, we will examine eukaryotic inheritance patterns that actually violate the laws of segregation or independent assortment.

5.1 OVERVIEW OF SIMPLE INHERITANCE PATTERNS

Before we delve more deeply into inheritance patterns, let's first compare a variety of inheritance patterns with each other (**Table 5.1**). We have already discussed two of these—simple dominant/recessive inheritance and X-linked inheritance—in Chapters 3 and 4, respectively. These various patterns occur because the outcome of a trait may be governed by two or more alleles in several different ways. Geneticists want to understand Mendelian inheritance from two perspectives. One goal is to predict the outcome of crosses. Many of the inheritance patterns described in Table 5.1 do not produce a 3:1 phenotypic ratio when two heterozygotes produce offspring. In this chapter, we consider how allelic interactions produce ratios that may differ from a simple Mendelian pattern. A second goal is to understand how the molecular expression of genes can account for an individual's phenotype. In other words, what is the underlying relationship between molecular genetics—the expression of genes to produce functional proteins—and the traits of individuals that inherit the genes? The remaining sections of this chapter will explore several patterns of inheritance from a molecular perspective.

5.1 REVIEWING THE KEY CONCEPTS

- Mendelian inheritance patterns obey Mendel's laws.
- Several inheritance patterns involving single genes differ from those observed by Mendel (see Table 5.1).

5.1 COMPREHENSION QUESTIONS

1. Which of the following statements is *false*?
   a. Mendelian inheritance patterns obey the law of segregation.
   b. Mendelian inheritance patterns obey the law of independent assortment.
   c. All inheritance patterns show a simple dominant/recessive relationship.
   d. None of the above are false.

TABLE **5.1**

Types of Mendelian Inheritance Patterns Involving Single Genes

Type	Description
Simple Mendelian	**Inheritance:** As described in Chapter 3, this term is commonly applied to the inheritance of alleles that obey Mendel's laws and follow a strict dominant/recessive relationship. In this chapter, we will see that some genes can be found in three or more alleles, making the relationship more complex. **Molecular:** The dominant allele encodes a functional protein, and 50% of the protein is sufficient to produce the dominant trait.
X-linked	**Inheritance:** As described in Chapter 4, this pattern involves the inheritance of genes that are located on the X chromosome. In mammals and fruit flies, males have a single copy of X-linked genes, whereas females have two copies. **Molecular:** If a pair of X-linked alleles shows a simple dominant/recessive relationship, the dominant allele encodes a functional protein, and 50% of the protein is sufficient to produce the dominant trait (in the heterozygous female).
Incomplete penetrance	**Inheritance:** This pattern occurs when a dominant phenotype is not expressed even though an individual carries a dominant allele. An example is an individual who carries the polydactyly allele but has a normal number of fingers and toes. **Molecular:** Even though a dominant gene may be present, the protein encoded by the gene may not exert its effects. This can be due to environmental influences or to other genes that may encode proteins that counteract the effects of the protein encoded by the dominant allele.
Incomplete dominance	**Inheritance:** This pattern occurs when the heterozygote has a phenotype that is intermediate between either corresponding homozygote. For example, a cross between homozygous red-flowered and homozygous white-flowered parents produces heterozygous offspring with pink flowers. **Molecular:** 50% of a functional protein is not sufficient to produce the same trait as the homozygote making 100%.
Overdominance	**Inheritance:** This pattern occurs when the heterozygote has a trait that is more beneficial than either homozygote. **Molecular:** Three common ways that heterozygotes gain benefits: (1) Their cells may have increased resistance to infection by microorganisms; (2) they may produce more forms of protein dimers, with enhanced function; or (3) they may produce proteins that function under a wider range of conditions.
Codominance	**Inheritance:** This pattern occurs when the heterozygote expresses both alleles simultaneously without forming an intermediate phenotype. For example, in blood typing, an individual carrying the *A* and *B* alleles has an AB blood type. **Molecular:** The codominant alleles encode proteins that function slightly differently from each other, and the function of each protein in the heterozygote affects the phenotype uniquely.
Sex-influenced inheritance	**Inheritance:** This pattern refers to the effect of sex on the phenotype of the individual. Some alleles are recessive in one sex and dominant in the opposite sex. An example is pattern baldness in humans. **Molecular:** Sex hormones may regulate the molecular expression of genes. This can influence the phenotypic effects of alleles.
Sex-limited inheritance	**Inheritance:** This refers to traits that occur in only one of the two sexes. An example is breast development in mammals. **Molecular:** Sex hormones may regulate the molecular expression of genes. This can influence the phenotypic effects of alleles. In this case, sex hormones that are primarily produced in only one sex are essential to produce a particular phenotype.
Lethal alleles	**Inheritance:** An allele that has the potential of causing the death of an organism. **Molecular:** Lethal alleles are most commonly loss-of-function alleles that encode proteins that are necessary for survival. In some cases, the allele may be due to a mutation in a nonessential gene that changes a protein to function with abnormal and detrimental consequences.

5.2 DOMINANT AND RECESSIVE ALLELES

In Chapters 3 and 4, we examined patterns of inheritance that showed a simple dominant/recessive relationship. This means that a heterozygote exhibits the dominant trait. In this section, we will take a closer look at why an allele may be dominant or recessive and discuss how dominant alleles may not always exert their effects.

Recessive Alleles Often Cause a Reduction in the Amount or Function of the Encoded Proteins

For any given gene, geneticists refer to prevalent alleles in a natural population as **wild-type alleles.** In large populations, more than one wild-type allele may occur—a phenomenon known as **genetic polymorphism.** For example, **Figure 5.1** illustrates a striking example of polymorphism in the elderflower orchid, *Dactylorhiza sambucina*. Throughout the range of this species in Europe, both yellow- and red-flowered individuals are prevalent. Both colors are considered wild type. At the molecular level, a wild-type allele typically encodes a protein that is made in the proper amount and functions normally. As discussed in Chapter 25, wild-type alleles tend to promote the reproductive success of organisms in their native environments.

In addition, random mutations occur in populations and alter preexisting alleles. Geneticists sometimes refer to these kinds of alleles as **mutant alleles** to distinguish them from the more common wild-type alleles. Because random mutations are more likely to disrupt gene function, mutant alleles are often defective in their ability to express a functional protein. Such mutant alleles tend to be rare in natural populations. They are typically, but not always, inherited in a recessive fashion.

Among Mendel's seven traits discussed in Chapter 3, the wild-type alleles are tall plants, purple flowers, axial flowers, yellow seeds, round seeds, green pods, and smooth pods (refer back to Figure 3.4). The mutant alleles are dwarf plants, white flowers, terminal flowers, green seeds, wrinkled seeds, yellow pods, and constricted pods. You may have already noticed that the seven wild-type alleles are dominant over the seven mutant alleles. Likewise, red eyes and normal wings are examples of wild-type alleles in *Drosophila*, and white eyes and miniature wings are recessive mutant alleles.

The idea that recessive alleles usually cause a substantial decrease in the expression of a functional protein is supported by the analysis of many human genetic diseases. Keep in mind that a genetic disease is usually caused by a mutant allele. **Table 5.2** lists several examples of human genetic diseases in which the

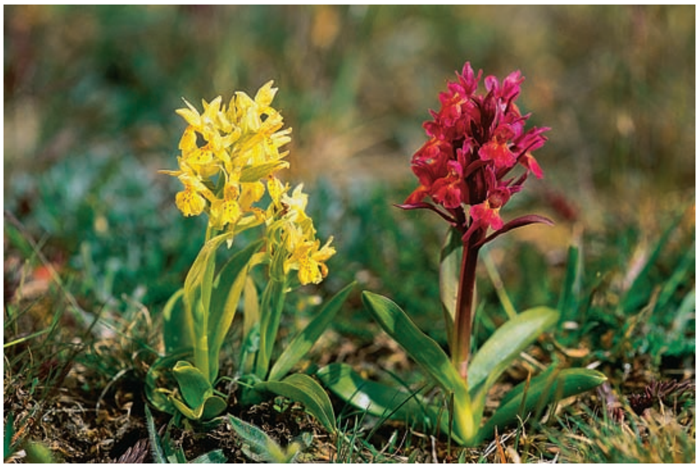


FIGURE 5.1 **An example of genetic polymorphism.** Both yellow and red flowers are common in natural populations of the elderflower orchid, *Dactylorhiza sambucina,* and both are considered wild type.

Concept check: Why are both of these colors considered to be wild type?

TABLE **5.2**

Examples of Recessive Human Diseases

Disease	Protein That Is Produced by the Normal Gene*	Description
Phenylketonuria	Phenylalanine hydroxylase	Inability to metabolize phenylalanine. The disease can be prevented by following a phenylalanine-free diet. If the diet is not followed early in life, it can lead to severe mental impairment and other symptoms.
Albinism	Tyrosinase	Lack of pigmentation in the skin, eyes, and hair.
Tay-Sachs disease	Hexosaminidase A	Defect in lipid metabolism. Leads to paralysis, blindness, and early death.
Sandhoff disease	Hexosaminidase B	Defect in lipid metabolism. Muscle weakness in infancy, early blindness, and progressive mental and motor deterioration.
Cystic fibrosis	Chloride transporter	Inability to regulate ion balance across epithelial cells. Leads to production of thick lung mucus and chronic lung infections.
Lesch-Nyhan syndrome	Hypoxanthine-guanine phosphoribosyl transferase	Inability to metabolize purines, which are bases found in DNA and RNA. Leads to self-mutilation behavior, poor motor skills, and usually mental impairment and kidney failure.

*Individuals who exhibit the disease are either homozygous for a recessive allele or hemizygous (for X-linked genes in human males). The disease symptoms result from a defect in the amount or function of the normal protein.

recessive allele fails to produce a specific cellular protein in its active form. In many cases, molecular techniques have enabled researchers to clone these genes and determine the differences between the wild-type and mutant alleles. They have found that the recessive allele usually contains a mutation that causes a defect in the synthesis of a fully functional protein.

To understand why many defective mutant alleles are inherited recessively, we need to take a quantitative look at protein function. With the exception of sex-linked genes, diploid individuals have two copies of every gene. In a simple dominant/recessive relationship, the recessive allele does not affect the phenotype of the heterozygote. In other words, a single copy of the dominant allele is sufficient to mask the effects of the recessive allele. If the recessive allele cannot produce a functional protein, how do we explain the wild-type phenotype of the heterozygote? As described in **Figure 5.2**, a common explanation is that 50% of the functional protein is adequate to provide the wild-type phenotype. In this example, the *PP* homozygote and *Pp* heterozygote each make sufficient functional protein to yield purple flowers. This means that the homozygous individual makes twice as much of the wild-type protein than it really needs to produce purple flowers. Therefore, if the amount is reduced to 50%, as in the heterozygote, the individual still has plenty of this protein to accomplish whatever cellular function it performs. The phenomenon that "50% of the normal protein is enough" is fairly common among many genes.

A second possible explanation for other genes is that the heterozygote actually produces more than 50% of the functional protein. Due to gene regulation, the expression of the normal gene may be increased, or "up-regulated," in the heterozygote to compensate for the lack of function of the defective allele. The topic of gene regulation is discussed in Chapters 16 and 17.

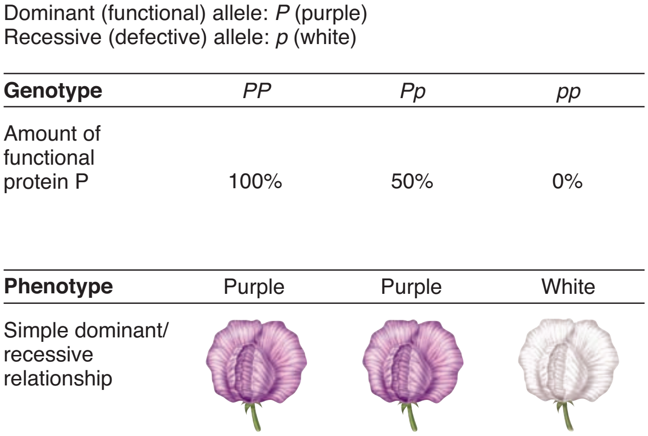

Dominant (functional) allele: P (purple)
Recessive (defective) allele: p (white)
Genotype
PP
Pp
pp
Amount of functional protein P
100%
50%
0%
Phenotype
Purple
Purple
White
Simple dominant/ recessive relationship

FIGURE 5.2 **A comparison of protein levels among homozygous and heterozygous genotypes *PP, Pp,* and *pp.***

Genes →Traits In a simple dominant/recessive relationship, 50% of the protein encoded by one copy of the dominant allele in the heterozygote is sufficient to produce the wild-type phenotype, in this case, purple flowers. A complete lack of the functional protein results in white flowers.

Concept check: Does a PP individual produce more of the protein encoded by the P gene than is necessary for the purple color?

Dominant Mutant Alleles Usually Exert Their Effects in One of Three Ways

Though dominant mutant alleles are much less common than recessive alleles, they do occur in natural populations. How can a mutant allele be dominant over a wild-type allele? Three explanations account for most dominant mutant alleles: a gain-of-function mutation, a dominant-negative mutation, or haploinsufficiency. Some dominant mutant alleles are due to **gain-of-function mutations.** Such mutations change the gene or the protein encoded by a gene so it gains a new or abnormal function. For example, a mutant gene may be overexpressed, thereby producing too much of the encoded protein. A second category are **dominant negative mutations** in which the protein encoded by the mutant gene acts antagonistically to the normal protein. In a heterozygote, the mutant protein counteracts the effects of the normal protein, thereby altering the phenotype. Finally, a third way that mutant alleles may affect phenotype is via **haploinsufficiency.** In this case, the mutant allele is a loss-of-function allele. Haploinsufficiency is used to describe patterns of inheritance in which a heterozygote (with one functional allele and one inactive allele) exhibits an abnormal or disease phenotype. An example in humans is a condition called polydactyly, in which a heterozygous individual has extra fingers or toes, which is discussed next.

Traits May Skip a Generation Due to Incomplete Penetrance and Vary in Their Expressivity

As we have seen, dominant alleles are expected to influence the outcome of a trait when they are present in heterozygotes. Occasionally, however, this may not occur. The phenomenon, called **incomplete penetrance,** is a situation in which an allele that is expected to cause a particular phenotype does not. **Figure 5.3a** illustrates a human pedigree for a dominant trait known as polydactyly. This trait causes the affected individual to have additional fingers or toes (or both) (**Figure 5.3b**). Polydactyly is due to an autosomal dominant allele—the allele is found in a gene located on an autosome (not a sex chromosome) and a single copy of this allele is sufficient to cause this condition. Sometimes, however, individuals carry the dominant allele but do not exhibit the trait. In Figure 5.3a, individual III-2 has inherited the polydactyly allele from his mother and passed the allele to a daughter and son. However, individual III-2 does not actually exhibit the trait himself, even though he is a heterozygote. In our polydactyly example, the dominant allele does not always "penetrate" into the phenotype of the individual. Alternatively, for recessive traits, incomplete penetrance occurs if a homozygote carrying the recessive allele does not exhibit the recessive trait. The measure of penetrance is described at the populational level. For example, if 60% of the heterozygotes carrying a dominant allele exhibit the trait, we would say that this trait is 60% penetrant. At the individual level, the trait is either present or not.

Another term used to describe the outcome of traits is the degree to which the trait is expressed, or its **expressivity.** In the case of polydactyly, the number of extra digits can vary. For example, one individual may have an extra toe on only one foot,

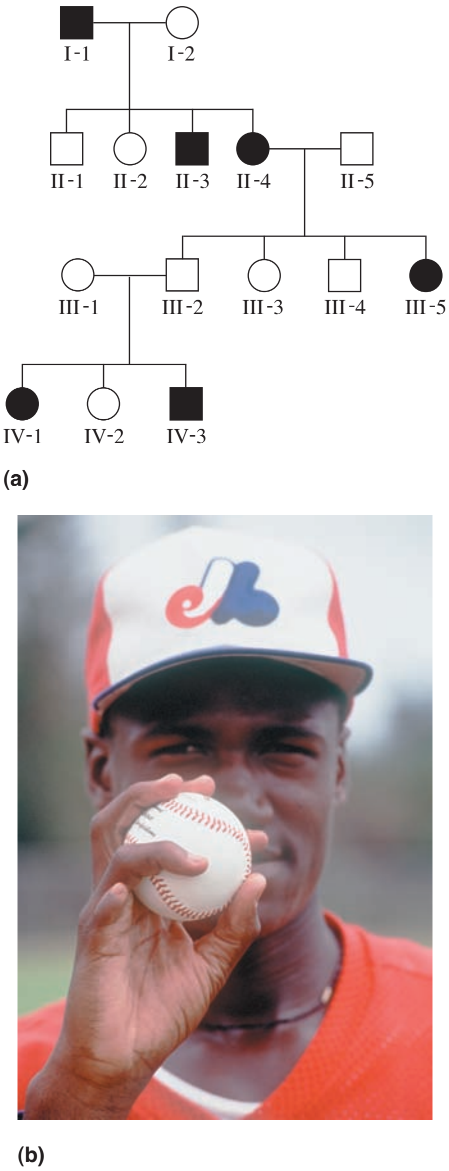


FIGURE 5.3 **Polydactyly, a dominant trait that shows incomplete penetrance.** **(a)** A family pedigree. Affected individuals are shown in black. Notice that offspring IV-1 and IV-3 have inherited the trait from a parent, III-2, who is heterozygous but does not exhibit polydactyly. **(b)** Antonio Alfonseca, a baseball player with polydactyly. His extra finger does not give him an advantage when pitching because it is small and does not touch the ball.

Concept check: Which individual(s) in this pedigree exhibit(s) incomplete penetrance?

whereas a second individual may have extra digits on both the hands and feet. Using genetic terminology, a person with several extra digits has a high expressivity of this trait, whereas a person with a single extra digit has a low expressivity.

How do we explain incomplete penetrance and variable expressivity? Although the answer may not always be understood, the range of phenotypes is often due to environmental influences and/or to effects of modifier genes in which one or more genes alter the phenotypic effects of another gene. We will consider the issue of the environment in the next section.

5.2 REVIEWING THE KEY CONCEPTS

- Wild-type alleles are defined as those that are prevalent in a population. When a gene exists in two or more wild-type alleles, this is an example of a genetic polymorphism (see Figure 5.1).
- Recessive alleles are often due to mutations that result in a reduction or loss of function of the encoded protein (see Figure 5.2 and Table 5.2).
- Dominant alleles are most commonly caused by gain-of-function mutations, dominant negative mutations, or haploinsufficiency.
- Incomplete penetrance is a situation in which an allele that is expected to be expressed is not expressed (see Figure 5.3).
- Traits may vary in their expressivity—the degree to which they are expressed.

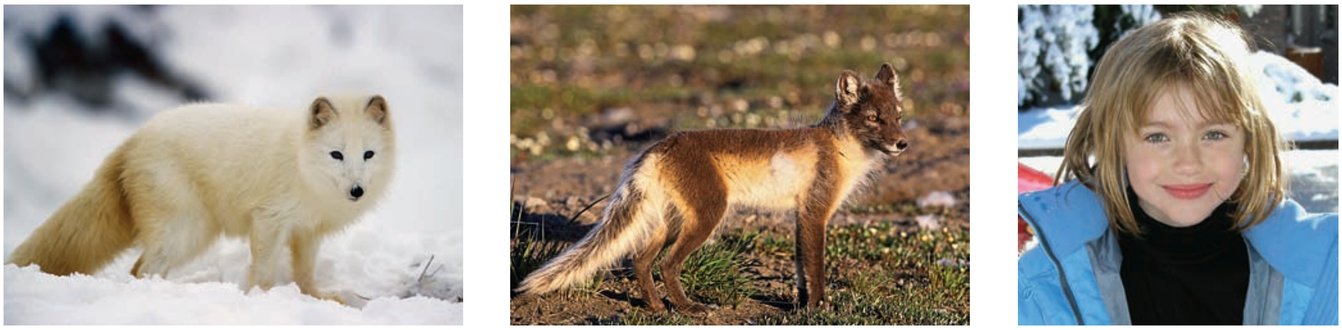
5.2 COMPREHENSION QUESTIONS

1. Which of the following is *not* an example of a wild-type allele?
   a. Yellow-flowered elderflower orchid
   b. Red-flowered elderflower orchid
   c. A gray elephant
   d. An albino (white) elephant
2. Dominant alleles may result from mutations that cause
   a. the overexpression of a gene or its protein product.
   b. a protein to inhibit the function of a normal protein.
   c. a protein to be inactive and 50% of the normal protein is insufficient for a normal phenotype.
   d. all of the above.
3. Polydactyly is a condition in which a person has extra fingers and/or toes. It is caused by a dominant allele. If a person carries this allele but does not have any extra fingers or toes, this is an example of
   a. haploinsufficiency.
   b. a dominant negative mutation.
   c. incomplete penetrance.
   d. a gain-of-function mutation.

5.3 ENVIRONMENTAL EFFECTS ON GENE EXPRESSION

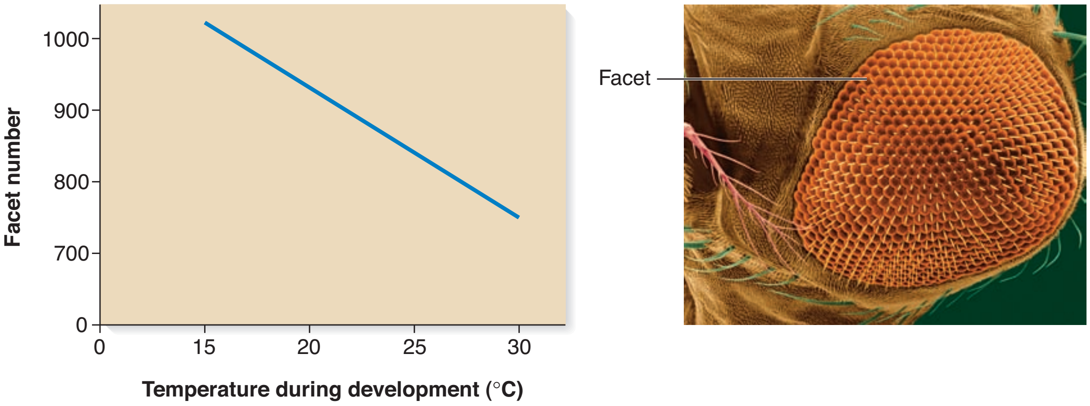
Throughout this book, our study of genetics tends to focus on the roles of genes in the outcome of traits. In addition to genetics, environmental conditions have a great effect on the phenotype of the individual. For example, the arctic fox (*Alopex lagopus*) goes through two color phases. During the cold winter, the arctic fox is primarily white, but in the warmer summer, it is mostly brown (**Figure 5.4a**). Such temperature-sensitive alleles affecting fur color are found among many species of mammals.

A dramatic example of the relationship between environment and phenotype can be seen in the human genetic disease known as phenylketonuria (PKU). This autosomal recessive disease is caused by a defect in a gene that encodes the enzyme phenylalanine hydroxylase. Homozygous individuals with this



(a) Arctic fox in winter and summer

(b) Healthy person with PKU



(c) Norm of reaction

FIGURE 5.4 **Variation in the expression of traits due to environmental effects. (a)** The arctic fox in the winter and summer. **(b)** A person with PKU who has followed a restricted diet and developed normally. **(c)** Norm of reaction. In this experiment, fertilized eggs from a population of genetically identical *Drosophila melanogaster* were allowed to develop into adult flies at different environmental temperatures. This graph shows the relationship between temperature (an environmental factor) and facet number in the eyes of the resulting adult flies. The micrograph shows the many facets of an eye of *D. melanogaster*.

Concept check: *What are the two main factors that determine an organism's traits?*

defective allele are unable to metabolize the amino acid phenylalanine properly. When given a standard diet containing phenylalanine, which is found in most protein-rich foods, PKU individuals manifest a variety of detrimental traits, including mental impairment, underdeveloped teeth, and foul-smelling urine. In contrast, when PKU individuals are diagnosed early and follow a restricted diet free of phenylalanine, they develop normally (**Figure 5.4b**). Since the 1960s, testing methods have been developed that can determine if an individual is lacking the phenylalanine hydroxylase enzyme. These tests permit the identification of infants who have PKU. Their diets can then be modified before the harmful effects of phenylalanine ingestion have occurred. As a result of government legislation, more than 90% of infants born in the United States are now tested for PKU. This test prevents a great deal of human suffering and is cost-effective. In the United States, the annual cost of PKU testing is estimated to be a few million dollars, whereas the cost of treating severely affected individuals with the disease would be hundreds of millions of dollars.

The examples of the arctic fox and PKU represent dramatic effects of very different environmental conditions. When considering the environment, geneticists often examine a range of conditions, rather than simply observing phenotypes under two different conditions. The term **norm of reaction** refers to the effects of environmental variation on a phenotype. Specifically, it is the phenotypic range seen in individuals with a particular genotype. To evaluate the norm of reaction, researchers begin with true-breeding strains that have the same genotypes and subject them to different environmental conditions. As an example, let's consider facet number in the eyes of fruit flies, *Drosophila melanogaster*. This species has compound eyes composed of many individual facets. **Figure 5.4c** shows the norm of reaction for facet number in genetically identical fruit flies that developed at different temperatures. As shown in the figure, the facet number varies with changes in temperature. At a higher temperature (30°C), the facet number is approximately 750, whereas at a lower temperature (15°C), it is over 1000.

5.3 REVIEWING THE KEY CONCEPTS

- The outcome of traits is also influenced by the environment (see Figure 5.4).

5.3 COMPREHENSION QUESTIONS

1. The outcome of an individual's traits is controlled by
   - a. genes.
   - b. the environment.
   - c. genes and the environment.
   - d. neither genes nor the environment.

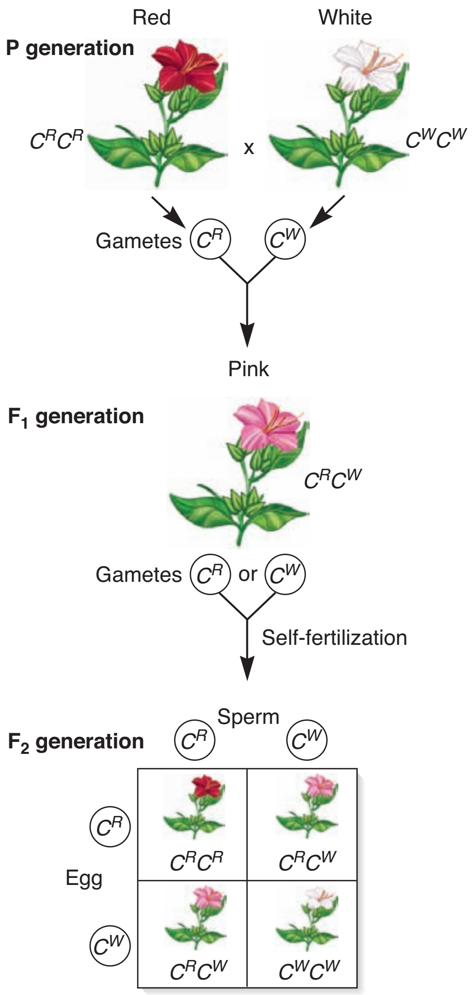
5.4 INCOMPLETE DOMINANCE, OVERDOMINANCE, AND CODOMINANCE

Thus far, we have considered inheritance patterns that follow a simple dominant/recessive inheritance pattern. In these cases, the heterozygote exhibits a phenotype that is the same as a homozygote that carries two copies of the dominant allele but different from the homozygote carrying two copies of the recessive allele. In this section, we will examine three different inheritance patterns in which the heterozygote shows a phenotype that is different from both types of homozygotes.

Incomplete Dominance Occurs When Two Alleles Produce an Intermediate Phenotype

Although many alleles display a simple dominant/recessive relationship, geneticists have also identified some cases in which a heterozygote exhibits **incomplete dominance**—a condition in which the phenotype is intermediate between the corresponding homozygous individuals. In 1905, the German botanist Carl Correns first observed this phenomenon in the color of the four-o'clock (*Mirabilis jalapa*). **Figure 5.5** describes Correns's experiment, in which a homozygous red-flowered four-o'clock plant was crossed to a homozygous white-flowered plant. The wild-type allele for red flower color is designated C^R and the white allele is C^W. As shown here, the offspring had pink flowers. If these F_1 offspring were allowed to self-fertilize, the F_2 generation consisted of 1/4 red-flowered plants, 1/2 pink-flowered plants, and 1/4 white-flowered plants. The pink plants in the F_2 generation were heterozygotes with an intermediate phenotype. As noted in the Punnett square in Figure 5.5, the F_2 generation displayed a 1:2:1 phenotypic ratio, which is different from the 3:1 ratio observed for simple Mendelian inheritance.

In Figure 5.5, incomplete dominance occurs because a heterozygote has an intermediate phenotype. At the molecular level, the allele that causes a white phenotype is expected to result in a lack of a functional protein required for pigmentation. Depending on the effects of gene regulation, the heterozygotes may produce only 50% of the normal protein, but this amount is not sufficient to produce the same phenotype as the C^RC^R homozygote, which may make twice as much of this protein. In this example, a reasonable explanation is that 50% of the functional protein cannot accomplish the same level of pigment synthesis that 100% of the protein can.



INTERACTIVE EXERCISE

FIGURE 5.5 **Incomplete dominance in the four-o'clock plant, *Mirabilis jalapa*.**

Genes →Traits When two different homozygotes (C^RC^R and C^WC^W) are crossed, the resulting heterozygote, C^RC^W, has an intermediate phenotype of pink flowers. In this case, 50% of the functional protein encoded by the C^R allele is not sufficient to produce a red phenotype.

Concept check: *At the molecular level, what is the explanation for why the flowers are pink instead of red?*

Finally, our opinion of whether a trait is dominant or incompletely dominant may depend on how closely we examine the trait in the individual. The more closely we look, the more likely we are to discover that the heterozygote is not quite the same as the wild-type homozygote. For example, Mendel studied the characteristic of pea seed shape and visually concluded that the *RR* and *Rr* genotypes produced round seeds and the *rr* genotype produced wrinkled seeds. The peculiar morphology of the wrinkled seed is caused by a large decrease in the amount of starch deposition in the seed due to a defective *r* allele. More recently, other scientists have dissected round and wrinkled seeds and examined their contents under the microscope. They have found that round seeds from heterozygotes actually contain an intermediate number of starch grains compared with seeds from the corresponding homozygotes (**Figure 5.6**). Within the seed, an intermediate amount of the functional protein is not enough to produce as many starch grains as in the homozygote carrying

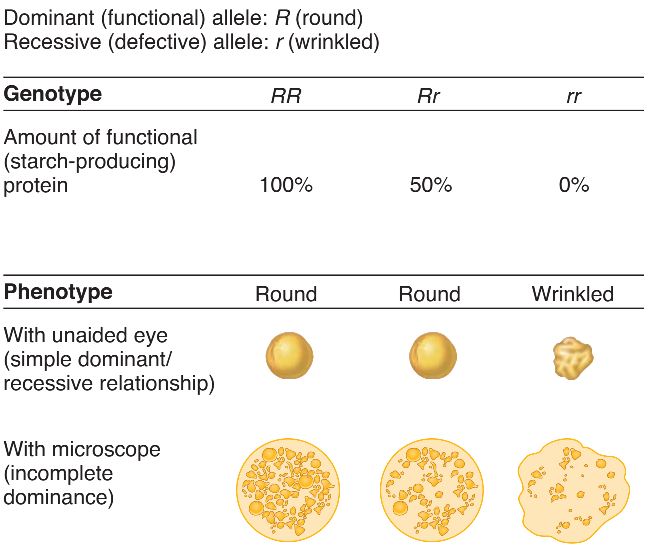

Dominant (functional) allele: R (round)
Recessive (defective) allele: r (wrinkled)
Genotype
RR
Rr
rr
Amount of functional (starch-producing) protein
100%
50%
0%
Phenotype
Round
Round
Wrinkled
With unaided eye (simple dominant/ recessive relationship)
With microscope (incomplete dominance)

FIGURE 5.6 **A comparison of phenotype at the macroscopic and microscopic levels.**

Genes →Traits This illustration shows the effects of a heterozygote having only 50% of the functional protein needed for starch production. This seed appears to be as round as those of the homozygote carrying the R allele, but when examined microscopically, it has produced only half the amount of starch.

Concept check: At which level is incomplete dominance more likely to be observed—at the molecular/cellular level or at the organism level?

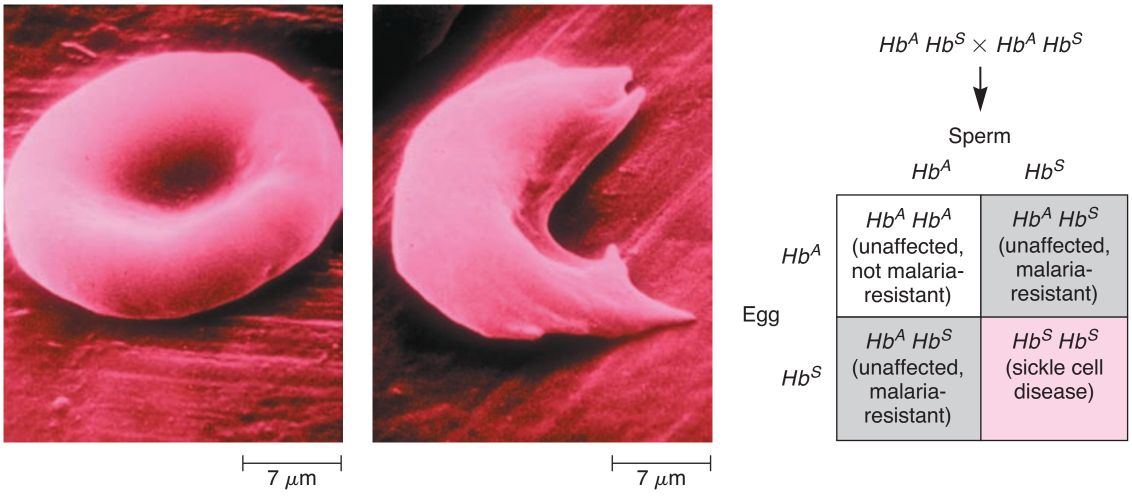
two copies of the R allele. Even so, at the level of our unaided eyes, heterozygotes produce seeds that appear to be round. With regard to phenotypes, the R allele is dominant to the r allele at the level of visual examination, but the R and r alleles show incomplete dominance at the level of starch biosynthesis.

Overdominance Occurs When Heterozygotes Have Superior Traits

As we have seen, the environment plays a key role in the outcome of traits. For certain genes, heterozygotes may display characteristics that are more beneficial for their survival in a particular environment. Such heterozygotes may be more likely to survive and reproduce. For example, a heterozygote may be larger, disease-resistant, or better able to withstand harsh environmental conditions. The phenomenon in which a heterozygote has greater reproductive success compared with either of the corresponding homozygotes is called **overdominance,** or **heterozygote advantage.**

A well-documented example involves a human allele that causes sickle cell disease in homozygous individuals. This disease is an autosomal recessive disorder in which the affected individual produces an altered form of the protein hemoglobin, which carries oxygen within red blood cells. Most people carry the Hb^A allele and make hemoglobin A. Individuals affected with sickle cell disease are homozygous for the Hb^S allele and produce only hemoglobin S. This causes their red blood cells to deform into a sickle shape under conditions of low oxygen concentration (**Figure 5.7a, b**). The sickling phenomenon causes the life span of these cells to be greatly shortened to only a few weeks compared with a normal span of 4 months, and therefore, anemia results. In addition, abnormal sickled cells can become clogged in the capillaries throughout the body, leading to localized areas of oxygen depletion. Such an event, called a crisis, causes pain and sometimes tissue and organ damage. For these reasons, the homozygous Hb^SHb^S individual usually has a shortened life span relative to an individual producing hemoglobin A.

In spite of the harmful consequences to homozygotes, the sickle cell allele has been found at a fairly high frequency among


7 μm
7 μm
$Hb^A Hb^S \times Hb^A Hb^S$
Sperm
Hb^A
Hb^S
Egg
Hb^A
Hb^S
$Hb^A Hb^A$ (unaffected, not malaria-resistant)
$Hb^A Hb^S$ (unaffected, malaria-resistant)
$Hb^A Hb^S$ (unaffected, malaria-resistant)
$Hb^S Hb^S$ (sickle cell disease)

(a) Normal red blood cell **(b) Sickled red blood cell** **(c) Example of sickle cell inheritance pattern**

INTERACTIVE EXERCISE

FIGURE 5.7 **Inheritance of sickle cell disease.** A comparison of **(a)** normal red blood cells and **(b)** those from a person with sickle cell disease. **(c)** The outcome of a cross between two heterozygous individuals.

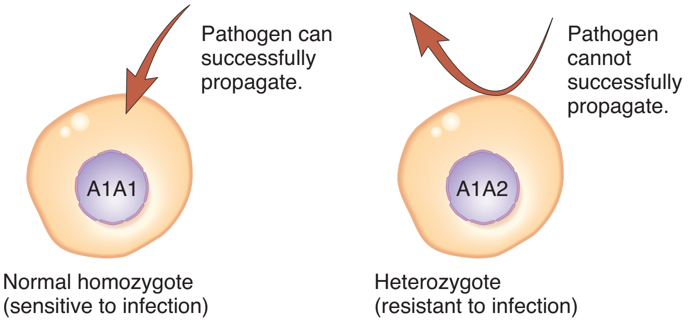
Concept check: Why does the heterozygote have an advantage?

human populations that are exposed to malaria. The protozoan genus that causes malaria, *Plasmodium*, spends part of its life cycle within the *Anopheles* mosquito and another part within the red blood cells of humans who have been bitten by an infected mosquito. However, red blood cells of heterozygotes, Hb^AHb^S, are likely to rupture when infected by this parasite, thereby preventing the parasite from propagating. People who are heterozygous have better resistance to malaria than do Hb^AHb^A homozygotes, while not incurring the ill effects of sickle cell disease. Therefore, even though the homozygous Hb^SHb^S condition is detrimental, the greater survival of the heterozygote has selected for the presence of the Hb^S allele within populations where malaria is prevalent. When viewing survival in such a region, overdominance explains the prevalence of the sickle cell allele. In Chapter 25, we will consider the role that natural selection plays in maintaining alleles that are beneficial to the heterozygote but harmful to the homozygote.

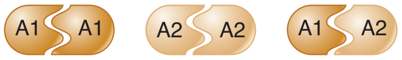
Figure 5.7c illustrates the predicted outcome when two heterozygotes have children. In this example, 1/4 of the offspring are Hb^AHb^A (unaffected, not malaria-resistant), 1/2 are Hb^AHb^S (unaffected, malaria-resistant), and 1/4 are Hb^SHb^S (sickle cell disease). This 1:2:1 ratio deviates from a simple Mendelian 3:1 phenotypic ratio.

Overdominance is usually due to two alleles that produce proteins with slightly different amino acid sequences. How can we explain the observation that two protein variants in the heterozygote produce a more favorable phenotype? There are three common explanations. The first is that the heterozygote has increased disease resistance. In the case of sickle cell disease, the phenotype is related to the infectivity of *Plasmodium* (**Figure 5.8a**). In the heterozygote, the infectious agent is less likely to propagate within red blood cells. Interestingly, researchers have speculated that other alleles in humans may confer disease resistance in the heterozygous condition but are detrimental in the homozygous state. These include PKU, in which the heterozygous fetus may be resistant to miscarriage caused by a fungal toxin, and Tay-Sachs disease, in which the heterozygote may be resistant to tuberculosis.

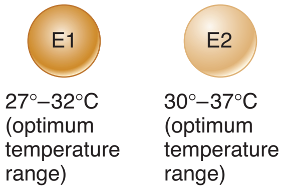
A second way to explain overdominance is related to the subunit composition of proteins. In some cases, a protein functions as a complex of multiple subunits; each subunit is composed of one polypeptide. A protein composed of two subunits is called a dimer. When both subunits are encoded by the same gene, the protein is a **homodimer.** The prefix *homo-* means that the subunits come from the same type of gene although the gene may exist in different alleles. **Figure 5.8b** considers a situation in which a gene exists in two alleles that encode polypeptides designated A1 and A2. Homozygous individuals can produce only A1A1 or A2A2 homodimers, whereas a heterozygote can also produce an A1A2 homodimer. Thus, heterozygotes produce three forms of the homodimer, homozygotes only one. For some proteins, A1A2 homodimers may have better functional activity because they are more stable or able to function under a wider range of conditions. The greater activity of the homodimer protein may be the underlying reason why a heterozygote has characteristics superior to either homozygote.


Pathogen can successfully propagate.
A1A1
Normal homozygote (sensitive to infection)
Pathogen cannot successfully propagate.
A1A2
Heterozygote (resistant to infection)

(a) Disease resistance



(b) Homodimer formation


E1
E2
27°–32°C (optimum temperature range)
30°–37°C (optimum temperature range)

(c) Variation in functional activity

FIGURE 5.8 **Three possible explanations for overdominance at the molecular level.**

Concept check: *Which of these three scenarios explains overdominance with regard to the sickle cell allele?*

A third molecular explanation of overdominance is that the proteins encoded by each allele exhibit differences in their functional activity. For example, let's suppose that a gene encodes a metabolic enzyme that can be found in two forms (corresponding to the two alleles), one that functions better at a lower temperature and the other that functions optimally at a higher temperature (**Figure 5.8c**). The heterozygote, which makes a mixture of both enzymes, may be at an advantage under a wider temperature range than either of the corresponding homozygotes.

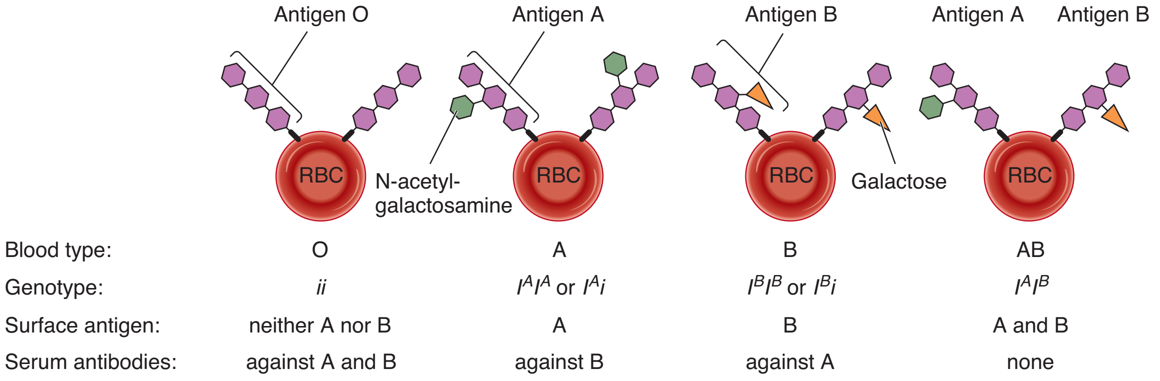
Alleles of the ABO Blood Group Can Be Dominant, Recessive, or Codominant

Thus far, we have considered examples in which a gene exists in two different alleles. As researchers have probed genes at the molecular level within natural populations of organisms, they have discovered that most genes exist in **multiple alleles.** Within a population, genes are typically found in three or more alleles. The ABO group of antigens, which determine blood type in humans, is an example of multiple alleles and illustrates an allelic relationship called codominance. To understand this concept, we first need to examine the molecular characteristics of human blood types. The plasma membranes of red blood cells have groups of interconnected sugars—oligosaccharides—that act as surface antigens (**Figure 5.9a**). Antigens are molecular structures that are recognized by antibodies produced by the immune system. On red blood cells, two different types of surface antigens, known as A and B may be found.

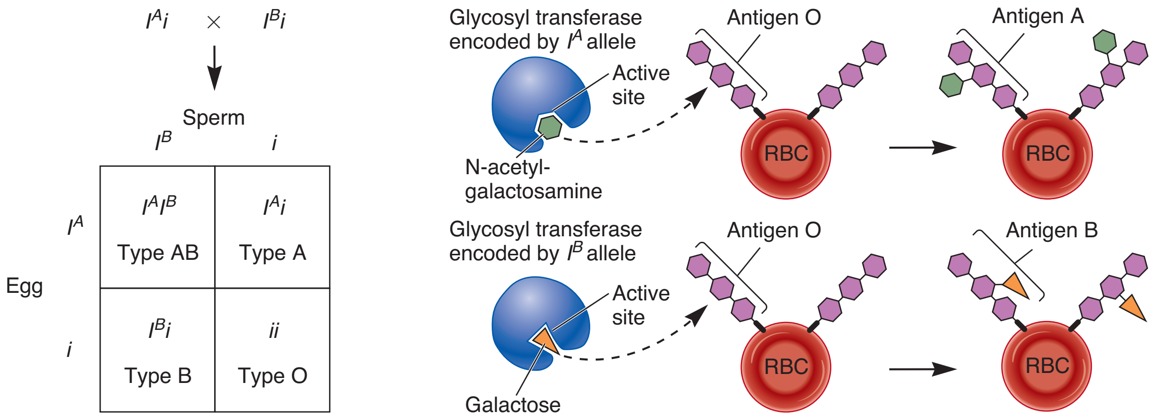
The synthesis of these surface antigens is controlled by two alleles, designated I^A and I^B, respectively. The i allele is recessive to both I^A and I^B. A person who is homozygous ii has type O blood and does not produce either antigen. A homozygous I^AI^A or heterozygous I^Ai individual has type A blood. The red blood cells of this individual contain the surface antigen known as A. Similarly, a homozygous I^BI^B or heterozygous I^Bi individual produces surface antigen B. As Figure 5.9a indicates, surface antigens A and B have significantly different molecular structures. A person who is I^AI^B has the blood type AB and expresses both surface antigens A and B. The phenomenon in which two alleles are both expressed in the heterozygous individual is called **codominance.** In this case, the I^A and I^B alleles are codominant to each other.

As an example of the inheritance of blood type, let's consider the possible offspring between two parents who are I^Ai and I^Bi (**Figure 5.9b**). The I^Ai parent makes I^A and i gametes, and the I^Bi parent makes I^B and i gametes. These combine to produce I^AI^B, I^Ai, I^Bi, and ii offspring in a 1:1:1:1 ratio. The resulting blood types are AB, A, B, and O, respectively.

Biochemists have analyzed the oligosaccharides produced on the surfaces of cells of differing blood types. In type O, the tree is smaller than type A or type B because a sugar has not been attached to a specific site on the tree. This idea is schematically shown in Figure 5.9a. How do we explain this difference at the molecular level? The gene that determines ABO blood type encodes an enzyme called glycosyl transferase that attaches a sugar to the oligosaccharide. The i allele carries a mutation that renders this enzyme inactive, which prevents the attachment of an additional sugar. By comparison, the two types of glycosyl transferase encoded by the I^A and I^B alleles have different structures in their active sites. The active site is the part of the protein that recognizes the sugar molecule that will be attached


Antigen O
Antigen A
Antigen B
Antigen A
Antigen B
RBC
N-acetyl-galactosamine
RBC
RBC
Galactose
RBC
Blood type: O A B AB
Genotype: ii I^AI^A or I^Ai I^BI^B or I^Bi I^AI^B
Surface antigen: neither A nor B A B A and B
Serum antibodies: against A and B against B against A none

(a) ABO blood type


I^Ai × I^Bi
Sperm
I^B
i
Egg
I^A
i
I^AI^B Type AB
I^Ai Type A
I^Bi Type B
ii Type O
Glycosyl transferase encoded by I^A allele
Active site
N-acetyl-galactosamine
Antigen O
RBC
Antigen A
RBC
Glycosyl transferase encoded by I^B allele
Active site
Galactose
Antigen O
RBC
Antigen B
RBC

(b) Example of the ABO inheritance pattern

(c) Formation of A and B antigen by glycosyl transferase

INTERACTIVE EXERCISE

FIGURE 5.9 **ABO blood type. (a)** A schematic representation of blood type at the cellular level. Note: This is not drawn to scale. A red blood cell is much larger than the oligosaccharide on the surface of the cell. **(b)** The predicted offspring from parents who are I^Ai and I^Bi. **(c)** The glycosyl transferase encoded by the I^A and I^B alleles recognizes different sugars due to changes in its active site. The i allele results in a nonfunctional enzyme.

Concept check: *Which allele is an example of a loss-of-function allele?*

to the oligosaccharide. The glycosyl transferase encoded by the I^A allele recognizes uridine diphosphate *N*-acetylgalactosamine (UDP-GalNAc) and attaches GalNAc to the oligosaccharide (**Figure 5.9c**). GalNAc is symbolized as a green hexagon. This produces the structure of surface antigen A. In contrast, the glycosyl transferase encoded by the I^B allele recognizes UDP-galactose and attaches galactose to the oligosaccharide. Galactose is symbolized as an orange triangle. This produces the molecular structure of surface antigen B. A person with type AB blood makes both types of enzymes and thereby has a tree with both types of sugar attached.

A small difference in the structure of the oligosaccharide, namely, a GalNAc in antigen A versus galactose in antigen B, explains why the two antigens are different from each other at the molecular level. These differences enable them to be recognized by different antibodies. A person who has blood type A makes antibodies to blood type B (refer back to Figure 5.9a). The antibodies against blood type B require a galactose in the oligosaccharide for their proper recognition. Their antibodies do not recognize and destroy their own blood cells, but they do recognize and destroy the blood cells from a type B person.

5.4 REVIEWING THE KEY CONCEPTS

- Incomplete dominance is an inheritance pattern in which the heterozygote has an intermediate phenotype (see Figure 5.5).
- Whether we judge an allele to be dominant or incompletely dominant may depend on how closely we examine the phenotype (see Figure 5.6).
- Overdominance is an inheritance pattern in which the heterozygote has greater reproductive success (see Figures 5.7, 5.8).
- Many genes exist in multiple alleles in a population.
- Some alleles, such as those that produce A and B blood antigens, are codominant (see Figure 5.9).

5.4 COMPREHENSION QUESTIONS

1. A pink-flowered four-o'clock is crossed to a red-flowered plant. What is the expected outcome for the offspring?
   a. All pink
   b. All red
   c. 1 red : 2 pink : 1 white
   d. 1 red : 1 pink
2. A person with type AB blood has a child with a person with type O blood. What are the possible blood types of the child?
   a. A and B
   b. A, B, and O
   c. A, B, AB, and O
   d. O only

5.5 SEX-INFLUENCED AND SEX-LIMITED INHERITANCE

In Chapter 4, we examined how the transmission pattern of X-linked genes depends on the sex of the parents and offspring. Sex can influence traits in other ways as well. The term **sex-influenced inheritance** refers to the phenomenon in which an allele is dominant in one sex but recessive in the opposite sex. Therefore, sex influence is a phenomenon of heterozygotes. Sex-influenced inheritance should not be confused with sex-linked inheritance. The genes that govern sex-influenced traits are almost always autosomal, which means they are not located on the X or Y chromosome.

In humans, the common form of pattern baldness provides an example of sex-influenced inheritance. As shown in **Figure 5.10**, the balding pattern is characterized by hair loss on the front and top of the head but not on the sides. This type of pattern baldness is inherited as an autosomal trait. (A common misconception is that this gene is X-linked.) When a male is heterozygous for the baldness allele, he will become bald.

Genotype	Phenotype	
	Males	Females
BB	Bald	Bald
Bb	Bald	Nonbald
bb	Nonbald	Nonbald

In contrast, a heterozygous female does not become bald. Women who are homozygous for the baldness allele develop the trait, but it is usually characterized by a significant thinning of the hair that occurs relatively late in life.

The sex-influenced nature of pattern baldness is related to the production of the male sex hormone testosterone. The gene that affects pattern baldness encodes an enzyme called 5-α-reductase, which converts testosterone to 5-α-dihydrotestosterone (DHT). DHT binds to cellular receptors and affects the expression of many genes, including those in the cells of the scalp. The allele that causes pattern baldness results in an overexpression of this enzyme. Because mature males normally make more testosterone than females, this allele has a greater phenotypic effect in males. However, a rare tumor of the adrenal gland can cause the secretion of abnormally large amounts of testosterone in females. If this occurs in a woman who is heterozygous *Bb*, she will become bald. If the tumor is removed surgically, her hair will return to its normal condition.

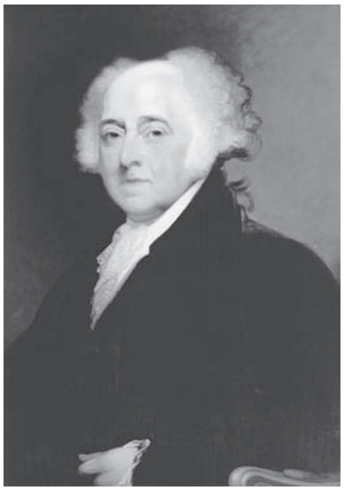


FIGURE 5.10 Pattern baldness in President John Adams.

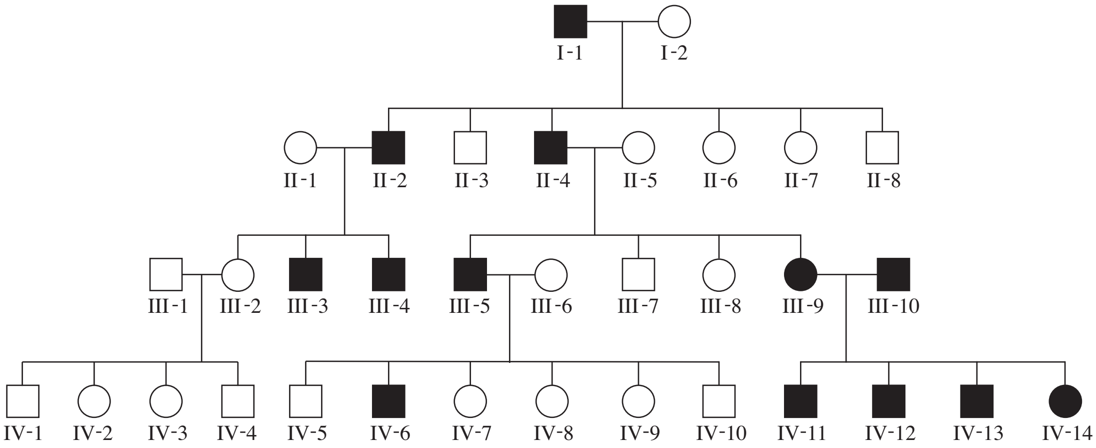
Concept check: Discuss whether John Adams inherited baldness from his mother or father or both.

The autosomal nature of pattern baldness has been revealed by the analysis of many human pedigrees. An example is shown in **Figure 5.11a.** A bald male may inherit the bald allele from either parent, and thus a striking observation is that bald fathers can pass this trait to their sons. This could not occur if the trait was X-linked, because fathers do not transmit an X chromosome to their sons. The analyses of many human pedigrees have shown that bald fathers, on average, have at least 50% bald sons. They are expected to produce an even higher percentage of bald male offspring if they are homozygous for the bald allele or if the mother also carries one or two copies of the bald allele. For example, a heterozygous bald male and heterozygous nonbald female produce 75% bald sons, and a homozygous bald male or homozygous bald female produce all bald sons.

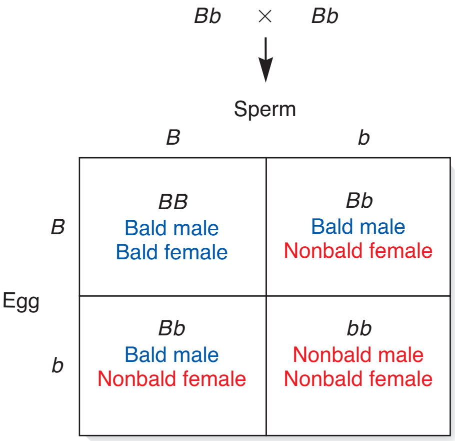
Figure 5.11b shows the predicted outcome if two heterozygotes produce offspring. In this Punnett square, the phenotypes are designated for both sons and daughters. *BB* offspring are bald, and *bb* offspring are not bald. *Bb* offspring are bald if they are sons and nonbald if they are daughters. The predicted genotypic ratio from this cross is 1 *BB* bald son to 1 *BB* bald daughter to 2 *Bb* bald sons to 2 *Bb* nonbald daughters to 1 *bb* nonbald son to 1 *bb* nonbald daughter. The predicted phenotypic ratio is 3 bald sons to 1 bald daughter to 3 nonbald daughters to 1 nonbald son. The ratio of bald to nonbald offspring is 4:4, which is the same as 1:1.

Another example in which sex affects an organism's phenotype is provided by **sex-limited inheritance,** in which a trait occurs in only one of the two sexes. The genes that influence sex-limited traits may be autosomal or sex-linked. In humans, examples of sex-limited traits are the presence of ovaries in females and the presence of testes in males. Due to these two sex-limited traits, mature females can only produce eggs, whereas mature males can only produce sperm.

Sex-limited traits are responsible for **sexual dimorphism** in which members of the opposite sex have different morphological features. This phenomenon is common among many animal species and is often striking among various species of birds in which the male has more ornate plumage than the female. As shown in **Figure 5.12**, roosters have a larger comb and wattles and longer neck and tail feathers than do hens. These sex-limited features are never in normal hens.



(a) A pedigree for human pattern baldness

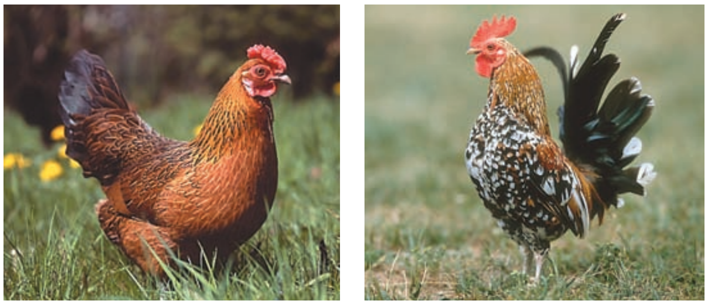


(b) Example of an inheritance pattern involving baldness

INTERACTIVE EXERCISE

FIGURE 5.11 **Inheritance of pattern baldness, a sex-influenced trait involving an autosomal gene. (a)** A family pedigree. Bald individuals are shown in black. **(b)** The predicted offspring from two heterozygous parents.

Concept check: *What is the phenotype of a woman who is heterozygous (one bald allele and one nonbald allele)?*



(a) Hen (b) Rooster

INTERACTIVE EXERCISE

FIGURE 5.12 **Differences in the feathering pattern in female and male chickens, an example of sex-limited inheritance.**

Concept check: What is the molecular explanation for sex-limited inheritance?

5.5 REVIEWING THE KEY CONCEPTS

- For sex-influenced traits such as pattern baldness in humans, heterozygous males and females have different phenotypes (see Figures 5.10, 5.11).
- Sex-limited traits are expressed in only one sex, thereby resulting in sexual dimorphism (see Figure 5.12).

5.5 COMPREHENSION QUESTIONS

1. A bald woman and nonbald man have a child. When this child grows up to be an adult, she or he could become
   a. a bald woman or a bald man.
   b. a nonbald woman or a bald man.
   c. a bald woman or a nonbald man.
   d. a nonbald woman or a nonbald man.

5.6 LETHAL ALLELES

Let's now turn our attention to alleles that have the most detrimental effect on phenotype—those that result in death. An allele that has the potential to cause the death of an organism is called a **lethal allele.** These are usually inherited in a recessive manner. When the absence of a specific protein results in a lethal phenotype, the gene that encodes the protein is considered an **essential gene** for survival. Though it varies according to species, researchers estimate that approximately 1/3 of all genes are essential genes. By comparison, **nonessential genes** are not absolutely required for survival, although they are likely to be beneficial to the organism. A loss-of-function mutation in a nonessential gene does not usually cause death. On rare occasions, however, a nonessential gene may acquire a mutation that causes the gene product to be abnormally expressed in a way that may interfere with normal cell function and lead to a lethal phenotype. Therefore, not all lethal mutations occur in essential genes, although the great majority do.

Many lethal alleles prevent cell division, thereby causing an organism to die at a very early stage. Others, however, may only exert their effects later in life. For example, a human genetic disease known as Huntington disease is caused by a dominant allele. The disease is characterized by a progressive degeneration of the nervous system, dementia, and early death. The age when these symptoms appear, or the **age of onset,** is usually between 30 and 50.

Other lethal alleles may kill an organism only when certain environmental conditions prevail. Such **conditional lethal alleles** have been extensively studied in experimental organisms. For example, some conditional lethals cause an organism to die only in a particular temperature range. These alleles, called **temperature-sensitive (ts) lethal alleles,** have been observed in many organisms, including *Drosophila.* A ts lethal allele may be fatal for a developing larva at a high temperature (30°C), but the larva survives if grown at a lower temperature (22°C). Temperature-sensitive lethal alleles are typically caused by mutations that alter the structure of the encoded protein so it does not function correctly at the nonpermissive temperature or becomes unfolded and is rapidly degraded. Conditional lethal alleles may also be identified when an individual is exposed to a particular agent in the environment. For example, people with a defect in the gene that encodes the enzyme glucose-6-phosphate dehydrogenase (G6PD) have a negative reaction to the ingestion of fava beans. This can lead to an acute hemolytic syndrome with a 10% mortality rate if not treated properly.

Finally, it is surprising that certain lethal alleles act only in some individuals. These are called **semilethal alleles.** Of course, any particular individual cannot be semidead. However, within a population, a semilethal allele will cause some individuals to die but not all of them. The reasons for semilethality are not always understood, but environmental conditions and the actions of other genes within the organism may help to prevent the detrimental effects of certain semilethal alleles. An example of a semilethal allele is the X-linked white-eyed allele, which was described in Chapter 4 (see Figure 4.9). Depending on the growth conditions, approximately 1/4 to 1/3 of the flies that would be expected to exhibit this white-eyed trait die prematurely.

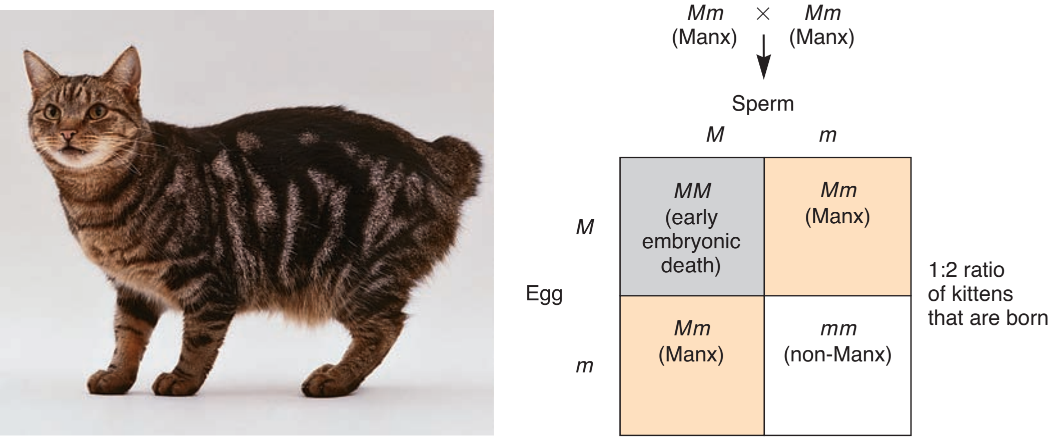
In some cases, a lethal allele may produce ratios that seemingly deviate from Mendelian ratios. An example is an allele in a breed of cats known as Manx, which originated on the Isle of Man (**Figure 5.13a**). The Manx cat carries a dominant mutation that affects the spine. This mutation shortens the tail, resulting in a range of tail lengths from normal to tailless. When two Manx cats are crossed to each other, the ratio of offspring is 1 normal to 2 Manx. How do we explain the 1:2 ratio? The answer is that about 1/4 of the offspring die during early embryonic development (**Figure 5.13b**). In this case, the Manx phenotype is dominant, whereas the lethal phenotype occurs only in the homozygous condition.

5.6 REVIEWING THE KEY CONCEPTS

- Lethal alleles most commonly occur in essential genes.
- Lethal alleles may result in inheritance patterns that yield unexpected ratios (see Figure 5.13).

5.6 COMPREHENSION QUESTIONS

1. The Manx phenotype in cats is caused by a dominant allele that is lethal in the homozygous state. A Manx cat is crossed to a normal



(a) A Manx cat

(b) Example of a Manx inheritance pattern

INTERACTIVE EXERCISE

FIGURE 5.13 **The Manx cat, which carries a lethal allele. (a)** Photo of a Manx cat, which typically has a shortened tail. **(b)** Outcome of a cross between two Manx cats. Animals that are homozygous for the dominant Manx allele (*M*) die during early embryonic development.

Concept check: Why do you think the heterozygote survives with developmental abnormalities, whereas the homozygote dies?

(non-Manx) cat. What is the expected outcome for the surviving offspring?

a. All Manx offspring
b. All normal offspring
c. 1 normal offspring : 1 Manx offspring
d. 1 normal offspring : 2 Manx offspring

5.7 PLEIOTROPY

Before ending our discussion of single-gene inheritance patterns, let's take a broader look at how a single gene may affect phenotype. Although we tend to discuss genes within the context of how they influence a single trait, most genes actually have multiple effects throughout a cell or throughout a multicellular organism. The multiple effects of a single gene on the phenotype of an organism is called **pleiotropy.** Pleiotropy occurs for several reasons, including the following:

1. The expression of a single gene can affect cell function in more than one way. For example, a defect in a microtubule protein may affect cell division and cell movement.
2. A gene may be expressed in different cell types in a multicellular organism. For example, a gene may be expressed in a muscle cell and also expressed in a nerve cell.
3. A gene may be expressed at different stages of development. For example, a gene may be expressed during embryonic development and also expressed in the adult.

In all or nearly all cases, the expression of a gene is pleiotropic with regard to the characteristics of an organism. The expression of any given gene influences the expression of many other genes in the genome, and vice versa. Pleiotropy is revealed when researchers study the effects of gene mutations. As an example of a pleiotropic mutation, let's consider cystic fibrosis, which is a recessive human disorder. In the late 1980s, the gene for cystic fibrosis was identified. It encodes a protein called the cystic fibrosis transmembrane conductance regulator (CFTR), which regulates ionic balance by allowing the transport of chloride ions (Cl^-) across epithelial cell membranes.

The mutation that causes cystic fibrosis diminishes the function of this Cl^- transporter, which affects several parts of the body in different ways. Because the movement of Cl^- affects water transport across membranes, the most severe symptom of cystic fibrosis is thick mucus in the lungs that occurs because of a water imbalance. In sweat glands, the normal Cl^- transporter has the function of recycling salt out of the glands and back into the skin before it can be lost to the outside world. Persons with cystic fibrosis have excessively salty sweat due to their inability to recycle salt back into their skin cells; a common test for cystic fibrosis is the measurement of salt on the skin. Another effect is seen in the reproductive system of males who are homozygous for the cystic fibrosis allele. Males with cystic fibrosis may be infertile because the vas deferens, the tubules that transport sperm from the testes, may be absent or undeveloped. Presumably, a normally functioning Cl^- transporter is needed for the proper development of the vas deferens in the embryo. Taken together, we can see that a defect in CFTR has multiple effects throughout the body.

5.7 REVIEWING THE KEY CONCEPTS

- Single genes usually exhibit pleiotropy, which means they exert multiple phenotypic effects.

5.7 COMPREHENSION QUESTIONS

1. Which of the following is a possible explanation for pleiotropy?
   a. The expression of a single gene can affect cell function in more than one way.

b. A gene may be expressed in different cell types in a multicellular organism.
c. A gene may be expressed at different stages of development.
d. All of the above are possible explanations.

5.8 GENE INTERACTIONS

Thus far, we considered the effects of a single gene on the outcome of a trait. This approach helps us to understand the various ways that alleles can influence traits. Researchers often examine the effects of a single gene on the outcome of a single trait as a way to simplify the genetic analysis. For example, Mendel studied one gene that affected the height of pea plants—tall versus dwarf alleles. Actually, many other genes in pea plants also affect height, but Mendel did not happen to study variants in those other height genes. How then did Mendel study the effects of a single gene? The answer lies in the genotypes of his strains. Although many genes affect the height of pea plants, Mendel chose true-breeding strains that differed with regard to only one of those genes. As a hypothetical example, let's suppose that pea plants have 10 genes affecting height, which we will call *K*, *L*, *M*, *N*, *O*, *P*, *Q*, *R*, *S*, and *T*. The genotypes of two hypothetical strains of pea plants may be

Tall strain:	*KK LL MM NN OO PP QQ RR SS TT*
Dwarf strain:	*KK LL MM NN OO PP QQ RR SS tt*

In this example, the alleles affecting height may differ at only a single gene. One strain is *TT* and the other is *tt*, and this accounts for the difference in their height. If we make crosses between these tall and dwarf strains, the genotypes of the F_2 offspring will differ with regard to only one gene; the other nine genes will be identical in all of them. This approach allows a researcher to study the effects of a single gene even though many genes may affect a single trait.

Researchers now appreciate that essentially all traits are affected by the contributions of many genes. Morphological features such as height, weight, growth rate, and pigmentation are all affected by the expression of many different genes in combination with environmental factors. In this section, we will further our understanding of genetics by considering how the allelic variants of two different genes affect a single trait. This phenomenon is known as **gene interaction.** Table 5.3 considers examples in which two different genes interact to influence the outcome of particular traits. In this section, we will examine some of these examples in greater detail.

TABLE 5.3

Types of Mendelian Inheritance Patterns Involving Two Genes

Type	Description
Epistasis	An inheritance pattern in which the alleles of one gene mask the phenotypic effects of the alleles of a different gene.
Complementation	A phenomenon in which two different parents that express the same or similar recessive phenotypes produce offspring with a wild-type phenotype.
Modifying genes	A phenomenon in which an allele of one gene modifies the phenotypic outcome of the alleles of a different gene.
Gene redundancy	A pattern in which the loss of function in a single gene has no phenotypic effect, but the loss of function of two genes has an effect. Functionality of only one of the two genes is necessary for a normal phenotype; the genes are functionally redundant.

A Cross Involving a Two-Gene Interaction Can Produce Two Distinct Phenotypes Due to Epistasis

William Bateson and Reginald Punnett discovered an unexpected gene interaction when studying crosses involving the sweet pea, *Lathyrus odoratus.* The wild sweet pea has purple flowers. However, they obtained several true-breeding mutant varieties with white flowers. Not surprisingly, when they crossed a true-breeding purple-flowered plant to a true-breeding white-flowered plant, the F_1 generation contained all purple-flowered plants and the F_2 generation (produced by self-fertilization of the F_1 generation) consisted of purple- and white-flowered plants in a 3:1 ratio.

A surprising result came in an experiment in which they crossed two different varieties of white-flowered plants (**Figure 5.14**). All of the F_1 generation plants had purple flowers! Bateson and Punnett then allowed the F_1 offspring to self-fertilize. The F_2 generation resulted in purple and white flowers in a ratio of 9 purple to 7 white. From this result, Bateson and Punnett deduced that two different genes were involved, with the following relationship:

C (one purple-color-producing) allele is dominant to *c* (white).

P (another purple-color-producing) allele is dominant to *p* (white).

cc or *pp* masks the *P* or *C* alleles, producing white color.

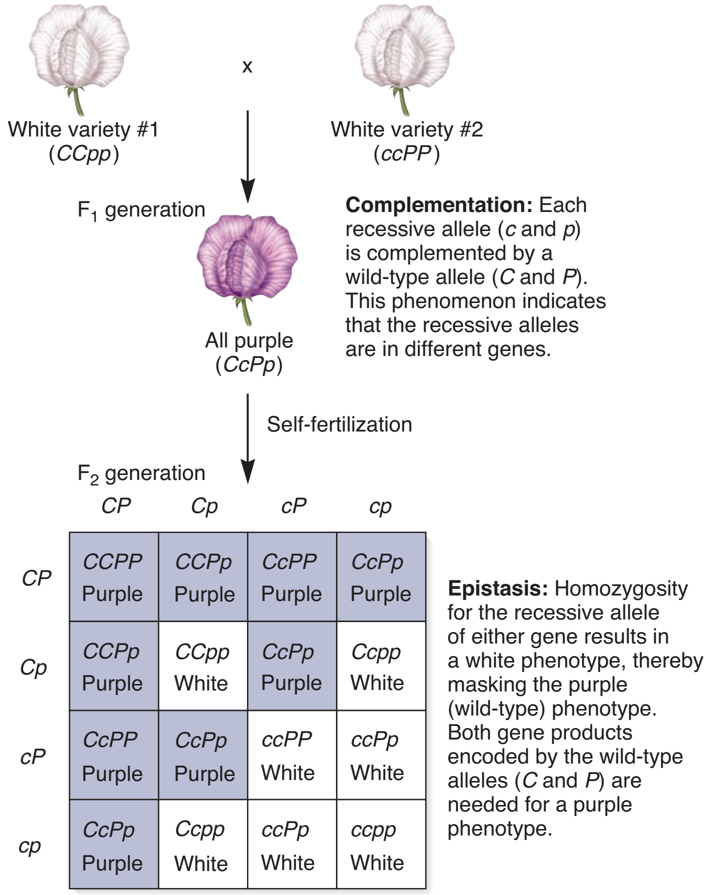
When the alleles of one gene mask the phenotypic effects of the alleles of another gene at a different locus, the phenomenon is called **epistasis.** Geneticists consider epistasis relative to a particular phenotype. If possible, geneticists use the wild-type phenotype as their reference phenotype when describing an epistatic interaction. In this case, purple flowers are wild type. Homozygosity for the white allele of one gene masks the expression of the purple-producing allele of another gene. In other words, the *cc* genotype is epistatic to a purple phenotype, and the *pp* genotype is also epistatic to a purple phenotype. At the level of genotypes, *cc* is epistatic to *PP* or *Pp*, and *pp* is epistatic to *CC* or *Cc*. This is an example of **recessive epistasis.** As seen in Figure 5.14, this epistatic interaction produces only two phenotypes—purple or white flowers—in a 9:7 ratio.

Epistasis often occurs because two (or more) different proteins participate in a common function. For example, two or more proteins may be part of an enzymatic pathway leading to the formation of a single product. To illustrate this idea, let's consider the formation of a purple pigment in the sweet pea.

$$\text{Colorless precursor} \xrightarrow{\text{Enzyme C}} \text{Colorless intermediate} \xrightarrow{\text{Enzyme P}} \textbf{Purple pigment}$$

In this example, a colorless precursor molecule must be acted on by two different enzymes to produce the purple pigment. Gene *C* encodes a functional protein called enzyme C, which converts the colorless precursor into a colorless intermediate. Two copies of the recessive allele (*cc*) result in a lack of production of this enzyme in the homozygote. Gene *P* encodes a functional enzyme P, which converts the colorless intermediate into the purple pigment. Like the *c* allele, the recessive *p* allele encodes a defective enzyme P. If an individual is homozygous for either recessive allele (*cc* or *pp*), it cannot make any functional enzyme C or enzyme P, respectively. When one of these enzymes is missing, purple pigment cannot be made, and the flowers remain white.

The parental cross shown in Figure 5.14 illustrates another genetic phenomenon called **complementation.** This term refers to the production of offspring with a wild-type phenotype from parents that both display the same or similar recessive phenotype. In this case, purple-flowered F_1 offspring were obtained from two white-flowered parents. Complementation typically occurs because the recessive phenotype in the parents is due to homozygosity at two different genes. In our sweet pea example, one parent is *CCpp* and the other is *ccPP*. In the F_1 offspring, the *C* and *P* alleles, which are wild-type and dominant, complement the *c* and *p* alleles, which are recessive. The offspring must have one wild-type allele of both genes to display the wild-type phenotype. Why is complementation an important experimental observation? When geneticists observe complementation in a genetic cross, the results suggest that the recessive phenotype in the two parent strains is caused by mutant alleles in two different genes.



Complementation: Each recessive allele (*c* and *p*) is complemented by a wild-type allele (*C* and *P*). This phenomenon indicates that the recessive alleles are in different genes.

	CP	*Cp*	*cP*	*cp*
CP	*CCPP* Purple	*CCPp* Purple	*CcPP* Purple	*CcPp* Purple
Cp	*CCPp* Purple	*CCpp* White	*CcPp* Purple	*Ccpp* White
cP	*CcPP* Purple	*CcPp* Purple	*ccPP* White	*ccPp* White
cp	*CcPp* Purple	*Ccpp* White	*ccPp* White	*ccpp* White

Epistasis: Homozygosity for the recessive allele of either gene results in a white phenotype, thereby masking the purple (wild-type) phenotype. Both gene products encoded by the wild-type alleles (*C* and *P*) are needed for a purple phenotype.

INTERACTIVE EXERCISE

FIGURE 5.14 A cross between two different white varieties of the sweet pea.

Concept check: What do the terms epistasis and complementation mean?

Due to Gene Redundancy, Loss-of-Function Alleles May Have No Effect on Phenotype

During the past several decades, researchers have discovered new kinds of gene interactions by studying model organisms such as *Escherichia coli* (a bacterium), *Saccharomyces cerevisiae* (baker's yeast), *Arabidopsis thaliana* (a model plant), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (a nematode worm), and *Mus musculus* (the laboratory mouse). The isolation of mutants that alter the phenotypes of these organisms has become a powerful tool for investigating gene function and has provided ways for researchers to identify new kinds of gene interactions. With the advent of modern molecular techniques (described in Chapters 18 to 20), a common approach for investigating gene function is to intentionally produce loss-of-function alleles in a gene of interest. When a geneticist abolishes gene function by creating an organism that is homozygous for a loss-of-function allele, the resulting organism is said to have undergone a **gene knockout.**

Why are gene knockouts useful? The primary reason for making a gene knockout is to understand how a gene affects the structure and function of cells or the phenotypes of organisms. For example, if a researcher knocked out a particular gene in a mouse and the resulting animal was unable to hear, the researcher would suspect that the role of the functional gene is to promote the formation of ear structures that are vital for hearing.

Interestingly, by studying gene knockouts in a variety of experimental organisms, geneticists have discovered that many knockouts have no obvious effect on phenotype at the cellular level or at the level of discernible traits. To explore gene function further, researchers may make two or more gene knockouts in the same organism. In some cases, gene knockouts in two different genes produce a phenotypic change even though the single knockouts have no effect (Figure 5.15). Geneticists may attribute this observation to **gene redundancy**—the phenomenon that one gene can compensate for the loss of function of another gene.

Gene redundancy may be due to different underlying causes. One common reason is gene duplication. Certain genes have been duplicated during evolution, so a species may contain two or more copies of similar genes. These copies, which are not identical due to the accumulation of random changes during evolution, are called **paralogs.** When one gene is missing, a

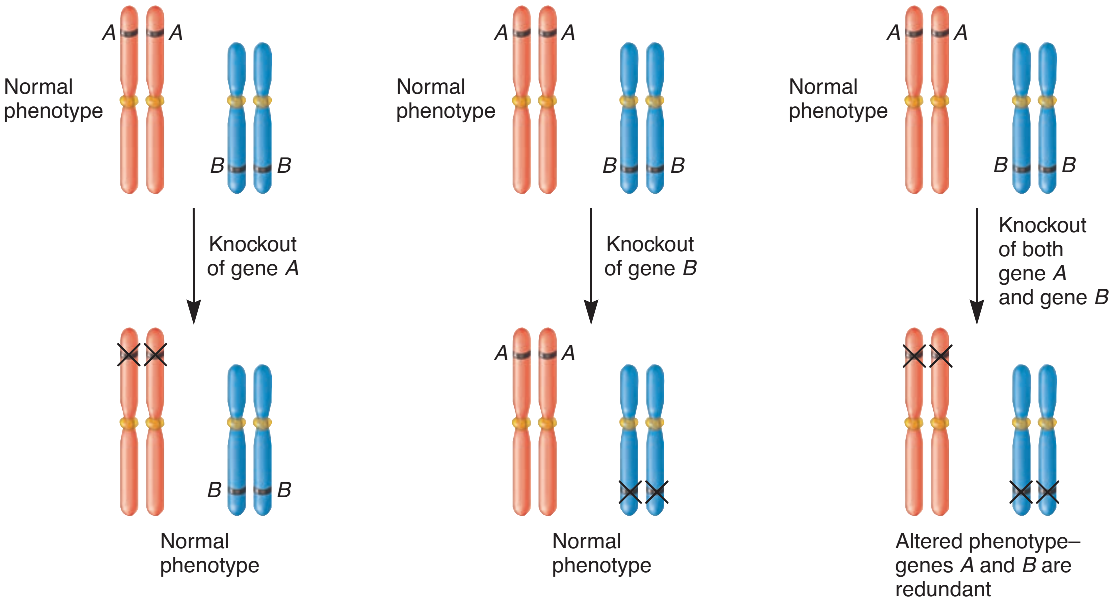

A
A
Normal phenotype
B
B
Knockout of gene *A*
B
B
Normal phenotype
A
A
Normal phenotype
B
B
Knockout of gene *B*
A
A
Normal phenotype
A
A
Normal phenotype
B
B
Knockout of both gene *A* and gene *B*
Altered phenotype–genes *A* and *B* are redundant

FIGURE 5.15 **A molecular explanation for gene redundancy.** To have a normal phenotype, an organism must have a functional copy of either gene *A* or gene *B*. If both gene *A* and gene *B* are knocked out, an altered phenotype occurs.

Concept check: *Explain why a single gene knockout does not have an effect on the phenotype?*

paralog may be able to carry out the missing function. For example, genes *A* and *B* in Figure 5.15 could be paralogs of each other. Alternatively, gene redundancy may involve proteins that are involved in a common cellular function. When one of the proteins is missing due to a gene knockout, the function of another protein may be increased to compensate for the missing protein and thereby overcome the defect.

Let's explore the consequences of gene redundancy in a genetic cross. George Shull conducted one of the first studies that illustrated the phenomenon of gene redundancy. His work involved *Capsella bursa-pastoris*, a weed known as shepherd's purse, a member of the mustard family. The trait he followed was the shape of the seed capsule, which is commonly triangular (**Figure 5.16**). Strains producing smaller ovate capsules are due to loss-of-function alleles in two different genes (*ttvv*). The ovate strain is an example of a double gene knockout. When Shull crossed a true-breeding plant with triangular capsules to a plant having ovate capsules, the F_1 generation all had triangular capsules. When the F_1 plants were self-fertilized, a surprising result came in the F_2 generation. Shull observed a 15:1 ratio of plants having triangular capsules to ovate capsules. The result can be explained by gene redundancy. Having at least one functional copy of either gene (*T* or *V*) is sufficient to produce the triangular phenotype. *T* and *V* are functional alleles of redundant genes.

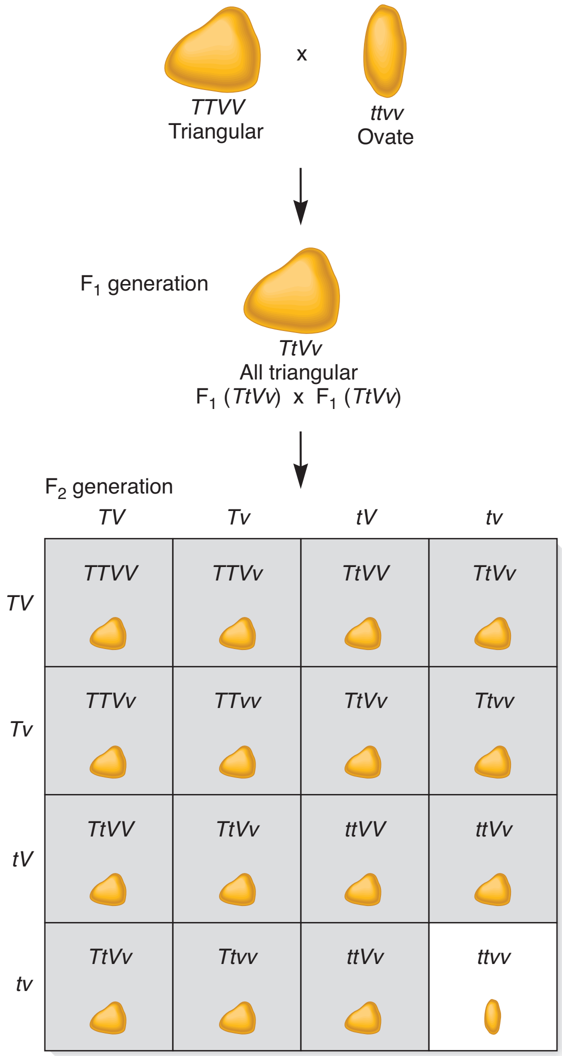

TTVV
Triangular
x
ttvv
Ovate
F1 generation
TtVv
All triangular
F1 (TtVv) x F1 (TtVv)
F2 generation
TV
Tv
tV
tv
TV
TTVV
TTVv
TtVV
TtVv
Tv
TTVv
TTvv
TtVv
Ttvv
tV
TtVV
TtVv
ttVV
ttVv
tv
TtVv
Ttvv
ttVv
ttvv

FIGURE 5.16 **Inheritance of capsule shape in *Capsella bursa-pastoris* (shepherd's purse), an example of gene redundancy.** In this case, triangular shape requires a dominant allele in one of two genes. The *T* and *V* alleles are redundant.

Concept check: *At the molecular level (with regard to loss-of-function alleles), explain why the ttvv homozygote has an ovate capsule phenotype.*

Only one of them is necessary for a triangular shape. When the functions of both genes are knocked out, as in the *ttvv* homozygote, the capsule becomes smaller and ovate.

5.8 REVIEWING THE KEY CONCEPTS

- A gene interaction is a situation in which two or more genes affect a single phenotype (see Table 5.3).
- Epistasis is a situation in which the allele of one gene masks the phenotypic expression of the alleles of a different gene. Complementation refers to the production of offspring with a wild-type phenotype from parents that both display the same or similar recessive phenotype (see Figure 5.14).
- Two different genes may have redundant functions, which is revealed in a double gene knockout (see Figures 5.15, 5.16).

5.8 COMPREHENSION QUESTIONS

1. Two different strains of sweet peas are true-breeding and have white flowers. When these strains are crossed together, the F_1 offspring all have purple flowers. This phenomenon is called
   a. epistasis.
   b. complementation.
   c. incomplete dominance.
   d. incomplete penetrance.
2. If the F_1 offspring from question 1 above are allowed to self-fertilize, what is the expected outcome for the F_2 offspring?
   a. all white
   b. all purple
   c. 3 purple to 1 white
   d. 9 purple to 7 white

KEY TERMS

Page 88. Mendelian inheritance, simple Mendelian inheritance
Page 90. wild-type alleles, genetic polymorphism, mutant alleles
Page 91. gain-of-function mutations, dominant negative mutations, haploinsufficiency, incomplete penetrance, expressivity
Page 93. norm of reaction
Page 94. incomplete dominance
Page 95. overdominance, heterozygote advantage
Page 96. homodimer, multiple alleles
Page 97. codominance
Page 98. sex-influenced inheritance
Page 99. sex-limited inheritance, sexual dimorphism
Page 100. lethal allele, essential gene, nonessential genes, age of onset, conditional lethal alleles, temperature-sensitive (ts) lethal alleles, semilethal alleles
Page 101. pleiotropy
Page 102. gene interaction, epistasis, recessive epistasis
Page 103. complementation, gene knockout, gene redundancy, paralogs

CHAPTER SUMMARY

- Mendelian inheritance patterns obey Mendel's laws.

5.1 Overview of Simple Inheritance Patterns

- Several inheritance patterns involving single genes differ from those observed by Mendel (see Table 5.1).

5.2 Dominant and Recessive Alleles

- Wild-type alleles are defined as those that are prevalent in a population. When a gene exists in two or more wild-type alleles, this is an example of a genetic polymorphism (see Figure 5.1).
- Recessive alleles are often due to mutations that result in a reduction or loss of function of the encoded protein (see Figure 5.2 and Table 5.2).
- Dominant alleles are most commonly caused by gain-of-function mutations, dominant negative mutations, or haploinsufficiency.
- Incomplete penetrance is a situation in which an allele that is expected to be expressed is not expressed (see Figure 5.3).
- Traits may vary in their expressivity.

5.3 Environmental Effects on Gene Expression

- The outcome of traits is also influenced by the environment (see Figure 5.4).

5.4 Incomplete Dominance, Overdominance, and Codominance

- Incomplete dominance is an inheritance pattern in which the heterozygote has an intermediate phenotype (see Figure 5.5).
- Whether we judge an allele to be dominant or incompletely dominant may depend on how closely we examine the phenotype (see Figure 5.6).
- Overdominance is an inheritance pattern in which the heterozygote has greater reproductive success (see Figures 5.7, 5.8).
- Many genes exist in multiple alleles in a population.
- Some alleles, such as those that produce A and B blood antigens, are codominant (see Figure 5.9).

5.5 Sex-Influenced and Sex-Limited Inheritance

- For sex-influenced traits such as pattern baldness in humans, heterozygous males and females have different phenotypes (see Figures 5.10, 5.11).
- Sex-limited traits are expressed in only one sex, thereby resulting in sexual dimorphism (see Figure 5.12).

5.6 Lethal Alleles

- Lethal alleles most commonly occur in essential genes. Lethal alleles may result in inheritance patterns that yield unexpected ratios (see Figure 5.13).

5.7 Pleiotropy

- Single genes usually exhibit pleiotropy, which means they exert multiple phenotypic effects.

5.8 Gene Interactions

- A gene interaction is a situation in which two or more genes affect a single phenotype (see Table 5.3).
- Epistasis is a situation in which the allele of one gene masks the phenotypic expression of the alleles of a different gene. Complementation refers to the production of offspring with a wild-type phenotype from parents that both display the same or similar recessive phenotype (see Figure 5.14).
- Two different genes may have redundant functions, which is revealed in a double gene knockout (see Figures 5.15, 5.16).

PROBLEM SETS & INSIGHTS

Solved Problems

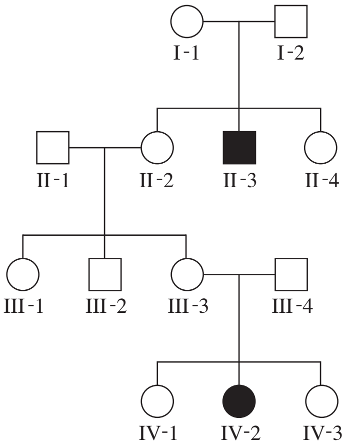
S1. In Ayrshire cattle, the spotting pattern of the animals can be either red and white or mahogany and white. The mahogany and white pattern is caused by the allele *M*. The red and white phenotype is controlled by the allele *m*. When mahogany and white animals are mated to red and white animals, the following results are obtained:

Genotype	**Phenotype**	
	Females	**Males**
MM	Mahogany and white	Mahogany and white
Mm	Red and white	Mahogany and white
mm	Red and white	Red and white

Explain the pattern of inheritance.

Answer: The inheritance pattern for this trait is sex-influenced inheritance. The *M* allele is dominant in males but recessive in females, whereas the *m* allele is dominant in females but recessive in males.

S2. The following pedigree involves a single gene causing an inherited disease. If you assume that incomplete penetrance is *not* occurring, indicate which modes of inheritance are possible. (Affected individuals are shown as filled symbols.)



A. Recessive

B. Dominant

C. X-linked, recessive

D. Sex-influenced, dominant in females

E. Sex-limited, recessive in females

Answer:

A. It could be recessive.

B. It could not be dominant (unless it was incompletely penetrant), because affected offspring have two unaffected parents.

C. It could not be X-linked recessive because individual IV-2 does not have an affected father.

D. It could not be sex-influenced dominant in females because individual II-3 (who would have to be homozygous) has an unaffected mother (who would have to be heterozygous and affected).

E. It could not be sex-limited because individual II-3 is an affected male and IV-2 is an affected female.

S3. Pattern baldness is an example of a sex-influenced trait that is dominant in males and recessive in females. A couple, neither of whom is bald, produced a bald son. What are the genotypes of the parents?

Answer: Because the father is not bald, we know he must be homozygous, *bb*. Otherwise, he would be bald. A female who is not bald can be either *Bb* or *bb*. Because she has produced a bald son, we know that she must be *Bb* in order to pass the *B* allele to her son.

S4. Two pink-flowered four-o'clocks were crossed to each other. What are the following probabilities for the offspring?

A. A plant will be red-flowered.

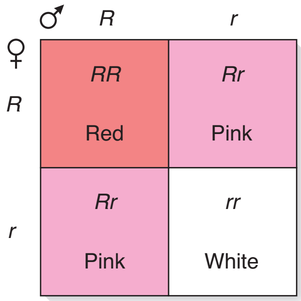
B. The first three plants examined will be white.

C. A plant will be either white or pink.

D. A group of six plants contain one pink, two whites, and three reds.

Answer: The first thing we need to do is construct a Punnett square to determine the individual probabilities for each type of offspring.

Because flower color is incompletely dominant, the cross is *Rr* × *Rr*.



♂ ♀	*R*	*r*
R	*RR* Red	*Rr* Pink
r	*Rr* Pink	*rr* White

The phenotypic ratio is 1 red to 2 pink to 1 white. In other words, 1/4 are expected to be red, 1/2 pink, and 1/4 white.

A. The probability of a red-flowered plant is 1/4, which equals 25%.

B. Use the product rule.

$1/4 \times 1/4 \times 1/4 = 1/64 = 1.6\%$

C. Use the sum rule because these are mutually exclusive events. A given plant cannot be both white and pink.

$1/4 + 1/2 = 3/4 = 75\%$

D. Use the multinomial expansion equation. See Solved Problem S7 in Chapter 3 for an explanation of the multinomial expansion equation. In this case, three phenotypes are possible.

$$P = \frac{n!}{a!b!c!} p^a q^b r^c$$

where

n = total number of offspring = 6
a = number of reds = 3
p = probability of reds = (1/4)
b = number of pinks = 1
q = probability of pink = (1/2)
c = number of whites = 2
r = probability of whites = (1/4)

If we substitute these values into the equation,

$$P = \frac{6!}{3!1!2!} (1/4)^3 (1/2)^1 (1/4)^2$$

$$P = 0.029 = 2.9\%$$

This means that 2.9% of the time we would expect to obtain six plants, three with red flowers, one with pink flowers, and two with white flowers.

Conceptual Questions

C1. Describe the differences among dominance, incomplete dominance, codominance, and overdominance.

C2. Discuss the differences among sex-influenced, sex-limited, and sex-linked inheritance. Describe examples.

C3. What is meant by a gene interaction? How can a gene interaction be explained at the molecular level?

C4. Let's suppose a recessive allele encodes a completely defective protein. If the functional allele is dominant, what does that tell you about the amount of the functional protein that is sufficient to cause the phenotype? What if the allele shows incomplete dominance?

C5. A nectarine is a peach without the fuzz. The difference is controlled by a single gene that is found in two alleles, *D* and *d*. At the molecular level, does it make more sense to you that the nectarine is homozygous for a recessive allele or that the peach is homozygous for the recessive allele? Explain your reasoning.

C6. An allele in *Drosophila* produces a "star-eye" trait in the heterozygous individual. (It is not X-linked.) However, the star-eye allele is lethal in homozygotes. What would be the ratio and phenotypes of surviving flies if star-eyed flies were crossed to each other?

C7. The serum from one individual (individual 1) is known to agglutinate the red blood cells from a second individual (individual 2). List the pairwise combinations of possible genotypes that individuals 1 and 2 could be. If individual 1 is the parent of individual 2, what are his or her possible genotypes?

C8. Which blood phenotypes (A, B, AB, and/or O) provide an unambiguous genotype? Is it possible for a couple to produce a family of children with all four blood types? If so, what must be the genotypes of the parents?

C9. A woman with type B blood has a child with type O blood. What are the possible genotypes and blood types of the father?

C10. A type A woman is the daughter of a type O father and type A mother. If she has children with a type AB man, what are the following probabilities?

A. A type AB child

B. A type O child

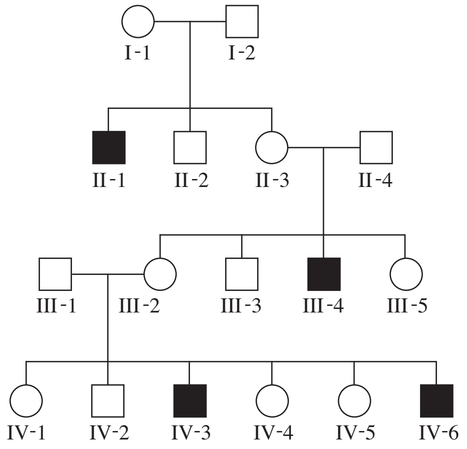
C. The first three children with type AB blood

D. A family containing two children with type B blood and one child with type AB

C11. In Shorthorn cattle, coat color is controlled by a single gene that can exist as a red allele (*R*) or white allele (*r*). The heterozygotes (*Rr*) have a color called roan that looks less red than the *RR* homozygotes. However, when examined carefully, the roan phenotype in cattle is actually due to a mixture of completely red hairs and completely white hairs. Should this be called incomplete dominance, codominance, or something else? Explain your reasoning.

C12. In chickens, the Leghorn variety has white feathers due to an autosomal dominant allele. Silkies have white feathers due to a recessive allele in a second (different) gene. If a true-breeding white Leghorn is crossed to a true-breeding white Silkie, what is the expected phenotype of the F_1 generation? If members of the F_1 generation are mated to each other, what is the expected phenotypic outcome of the F_2 generation? Assume the chickens in the parental generation are homozygous for the white allele at one gene and homozygous for the brown allele at the other gene. In subsequent generations, nonwhite birds will be brown.

C13. Propose the most likely mode of inheritance (autosomal dominant, autosomal recessive, or X-linked recessive) for the following pedigree.



Affected individuals are shown with filled (black) symbols.

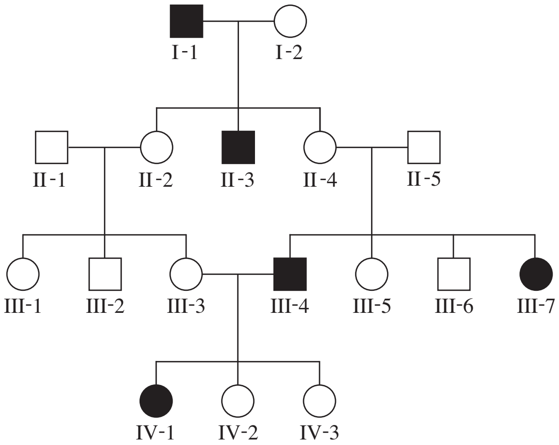
C14. With regard to pattern baldness in humans (a sex-influenced trait), a woman who is not bald and whose mother is bald has children with a bald man whose father is not bald. What are their probabilities of having the following types of families?

A. Their first child will not become bald.

B. Their first child will be a male who will not become bald.

C. Their first three children will be females who are not bald.

C15. In the pedigree shown here for a trait determined by a single gene (affected individuals are shown in black), state whether it would be possible for the trait to be inherited in each of the following ways:



A. Recessive

B. X-linked recessive

C. Dominant, complete penetrance

D. Sex-influenced, dominant in males

E. Sex-limited

F. Dominant, incomplete penetrance

C16. Let's suppose you have pedigree data from thousands of different families involving a particular genetic disease. How would you decide whether the disease is inherited as a recessive trait as opposed to one that is dominant with incomplete penetrance?

C17. Compare phenotypes at the molecular, cellular, and organism levels for individuals who are homozygous for the hemoglobin allele, Hb^AHb^A, and the sickle cell allele, Hb^SHb^S.

C18. A very rare dominant allele that causes the little finger to be crooked has a penetrance value of 80%. In other words, 80% of heterozygotes carrying the allele have a crooked little finger. If a homozygous unaffected person has children with a heterozygote carrying this mutant allele, what is the probability that an offspring will have little fingers that are crooked?

C19. A sex-influenced trait in humans is one that affects the length of the index finger. A "short" allele is dominant in males and recessive in females. Heterozygous males have an index finger that is significantly shorter than the ring finger. The gene affecting index finger length is located on an autosome. A woman with short index fingers has children with a man who has normal index fingers. They produce five children in the following order: female, male, male, female, male. The oldest female offspring marries a man with normal fingers and then has one daughter. The youngest male among the five children marries a woman with short index fingers, and then they have two sons. Draw the pedigree for this family. Indicate the phenotypes of every individual (filled symbols for individuals with short index fingers and open symbols for individuals with normal index fingers).

C20. In horses, there are three coat-color patterns termed cremello (beige), chestnut (brown), and palomino (golden with light mane and tail). If two palomino horses are mated, they produce about 1/4 cremello, 1/4 chestnut, and 1/2 palomino offspring. In contrast, cremello horses and chestnut horses breed true. (In other words, two cremello horses will produce only cremello offspring and two chestnut horses will produce only chestnut offspring.) Explain this pattern of inheritance.

Application and Experimental Questions

E1. A seed dealer wants to sell four-o'clock seeds that will produce only red, white, or pink flowers. Explain how this should be done.

E2. Mexican hairless dogs have little hair and few teeth. When a Mexican hairless is mated to another breed of dog, about half of the puppies are hairless. When two Mexican hairless dogs are mated to each other, about 1/3 of the surviving puppies have hair, and about 2/3 of the surviving puppies are hairless. However, about two out of eight puppies from this type of cross are born grossly deformed and do not survive. Explain this pattern of inheritance.

E3. In chickens, some varieties have feathered shanks (legs), but others do not. In a cross between a Black Langhans (feathered shanks) and Buff Rocks (unfeathered shanks), the shanks of the F_1 generation are all feathered. When the F_1 generation is crossed, the F_2 generation contains chickens with feathered shanks to unfeathered shanks in a ratio of 15:1. Suggest an explanation for this result.

E4. In sheep, the formation of horns is a sex-influenced trait; the allele that results in horns is dominant in males and recessive in females. Females must be homozygous for the horned allele to have horns. A horned ram was crossed to a polled (unhorned) ewe, and the first offspring they produced was a horned ewe. What are the genotypes of the parents?

E5. A particular breed of dog can have long hair or short hair. When true-breeding long-haired animals were crossed to true-breeding short-haired animals, the offspring all had long hair. The F_2 generation produced a 3:1 ratio of long- to short-haired offspring. A second gene affects the texture of the hair. The two variants are wiry hair and straight hair. F_1 offspring from a cross of these two varieties all had wiry hair, and F_2 offspring showed a 3:1 ratio of wiry- to straight-haired puppies. Recently, a breeder of the short-, wiry-haired dogs found a female puppy that was albino. Similarly, another breeder of the long-, straight-haired dogs found a male puppy that was albino. Because the albino trait is always due to a recessive allele, the two breeders got together and mated the two dogs. Surprisingly, all the puppies in the litter had black hair. How would you explain this result?

E6. In the clover butterfly, males are always yellow, but females can be yellow or white. In females, white is a dominant allele. Two yellow butterflies were crossed to yield an F_1 generation consisting of 50% yellow males, 25% yellow females, and 25% white females. Describe how this trait is inherited and the genotypes of the parents.

E7. Duroc Jersey pigs are typically red, but a sandy variation is also seen. When two different varieties of true-breeding sandy pigs were crossed to each other, they produced F_1 offspring that were red. When these F_1 offspring were crossed to each other, they produced red, sandy, and white pigs in a 9:6:1 ratio. Explain this pattern of inheritance.

E8. In certain species of summer squash, fruit color is determined by two interacting genes. A dominant allele, *W*, determines white color, and a recessive allele (*w*) allows the fruit to be colored. In a homozygous *ww* individual, a second gene determines fruit color: *G* (green) is dominant to *g* (yellow). A white squash and a yellow squash were crossed, and the F_1 generation yielded approximately 50% white fruit and 50% green fruit. What are the genotypes of the parents?

E9. Certain species of summer squash exist in long, spherical, or disk shapes. When a true-breeding long-shaped strain was crossed to a true-breeding disk-shaped strain, all of the F_1 offspring were disk-shaped. When the F_1 offspring were allowed to self-fertilize, the F_2 generation consisted of a ratio of 9 disk-shaped to 6 round-shaped to 1 long-shaped. Assuming the shape of summer squash is governed by two different genes, with each gene existing in two alleles, propose a mechanism to account for this 9:6:1 ratio.

E10. In a species of plant, two genes control flower color. The red allele (*R*) is dominant to the white allele (*r*); the color-producing allele (*C*) is dominant to the non-color-producing allele (*c*). You suspect that either an *rr* homozygote or a *cc* homozygote will produce white flowers. In other words, *rr* is epistatic to *C*, and *cc* is epistatic to *R*. To test your hypothesis, you allowed heterozygous plants (*RrCc*) to self-fertilize and counted the offspring. You obtained the following data: 201 plants with red flowers and 144 with white flowers. Conduct a chi-square analysis to see if your observed data are consistent with your hypothesis.

E11. In *Drosophila*, red eyes is the wild-type phenotype. Several different genes (with each gene existing in two or more alleles) are known to affect eye color. One allele causes purple eyes, and a different allele causes sepia eyes. Both of these alleles are recessive compared with red eye color. When flies with purple eyes were crossed to flies with sepia eyes, all of the F_1 offspring had red eyes. When the F_1 offspring were allowed to mate with each other, the following data were obtained:

146 purple eyes

151 sepia eyes

50 purplish sepia eyes

444 red eyes

Explain this pattern of inheritance. Conduct a chi-square analysis to see if the experimental data fit your hypothesis.

Questions for Student Discussion/Collaboration

1. Let's suppose a gene exists as a functional wild-type allele and a nonfunctional mutant allele. At the organism level, the wild-type allele is dominant. In a heterozygote, discuss whether dominance occurs at the cellular or molecular level. Discuss examples in which the issue of dominance depends on the level of examination.
2. In oats, the color of the chaff is determined by a two-gene interaction. When a true-breeding black plant was crossed to a true-breeding white plant, the F_1 generation was composed of all black plants. When the F_1 offspring were crossed to each other, the ratio produced was 12 black to 3 gray to 1 white. First, construct a Punnett square that accounts for this pattern of inheritance. Which genotypes produce the gray phenotype? Second, at the level of protein function, how would you explain this type of inheritance?

Answers to Comprehension Questions

5.1: c

5.2: d, d, c

5.3: c

5.4: d, a

5.5: b

5.6: c

5.7: d

5.8: b, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

6

CHAPTER OUTLINE

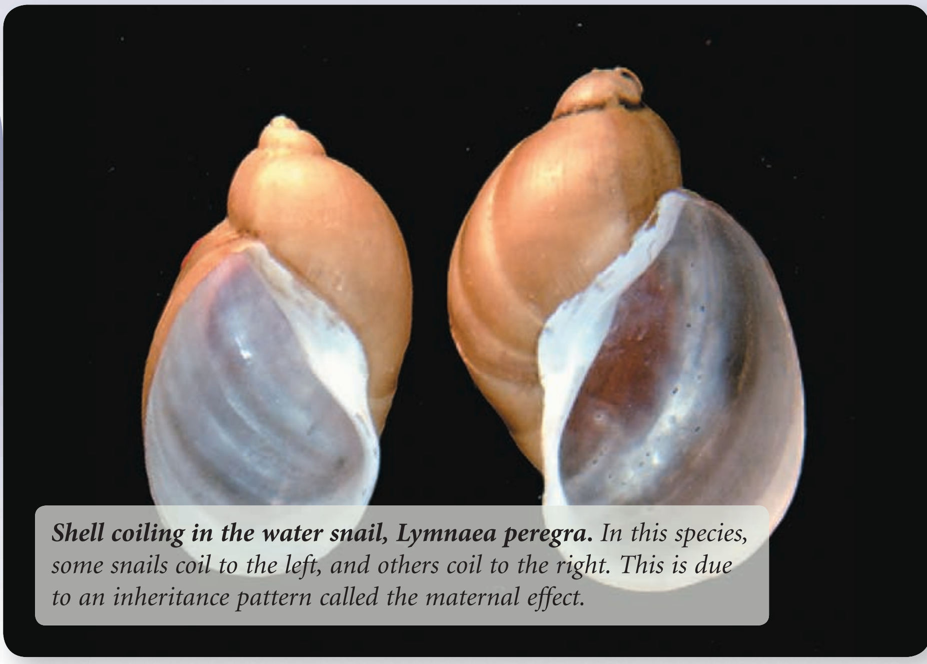
6.1 Extranuclear Inheritance: Chloroplasts

6.2 Extranuclear Inheritance: Mitochondria

6.3 Theory of Endosymbiosis

6.4 Epigenetics: Imprinting

6.5 Maternal Effect



Shell coiling in the water snail, Lymnaea peregra. *In this species, some snails coil to the left, and others coil to the right. This is due to an inheritance pattern called the maternal effect.*

EXTRANUCLEAR INHERITANCE, IMPRINTING, AND MATERNAL EFFECT

Mendelian inheritance patterns involve genes that directly influence the outcome of an offspring's traits and obey Mendel's laws. To predict phenotype, we must consider several factors. These include the dominant/recessive relationship of alleles, gene interactions that may affect the expression of a single trait, and the roles that sex and the environment play in influencing the individual's phenotype. Once these factors are understood, we can predict the phenotypes of offspring from their genotypes.

Most genes in eukaryotic species follow a Mendelian pattern of inheritance. However, many genes do not. In this chapter, we will examine several additional and even bizarre types of inheritance patterns that deviate from a Mendelian pattern. In the first two sections, we will examine deviations from Mendelian inheritance that arise because some genetic material is not located in the cell nucleus. Certain cellular organelles, such as chloroplasts and mitochondria, contain their own genetic material. We will survey the inheritance of these organellar genes. We then consider two interesting examples of non-Mendelian inheritance called imprinting and the maternal effect. Even though these inheritance patterns involve genes that are on chromosomes within the cell nucleus, the genotype of the offspring does not directly govern their phenotype in ways predicted by Mendel. We will see how gene silencing and the timing of gene expression can cause a non-Mendelian pattern of inheritance.

6.1 EXTRANUCLEAR INHERITANCE: CHLOROPLASTS

One cause of non-Mendelian inheritance patterns involves genes that are not located in the cell nucleus. In eukaryotic species, the most biologically important example of extranuclear inheritance is due to genetic material in cellular organelles. In addition to the cell nucleus, the chloroplasts and mitochondria contain their own DNA. Because these organelles are found within the cytoplasm of the cells, the inheritance of organellar genetic material is called **extranuclear inheritance** (the prefix *extra-* means outside of), or **cytoplasmic inheritance.** In this section, we will examine the genetic composition of chloroplasts and explore the patterns of transmission of this organelle from parent to offspring.

Chloroplasts Contain Circular Chromosomes with Many Genes

In 1951, Yukako Chiba was the first to suggest that chloroplasts contain their own DNA. He based his conclusion on the staining properties of a DNA-specific dye known as Feulgen. Researchers later developed techniques to purify organellar DNA. In addition, electron microscopy studies provided interesting insights into the organization and composition of chloroplast (and mitochondrial) chromosomes. More recently, the advent of molecular genetic techniques in the 1970s and 1980s has allowed researchers to determine the sequences of organellar DNAs. From these types of studies, the chromosomes of chloroplasts were found to resemble smaller versions of bacterial chromosomes.

The genetic material of chloroplasts is located inside the organelle in a region known as the **nucleoid** (**Figure 6.1**). The genome is a single circular chromosome (composed of double-stranded DNA), although a nucleoid may contain multiple copies of this chromosome. In addition, a chloroplast often has more than one nucleoid. **Table 6.1** describes the genetic composition of chloroplasts for a few selected species.

Among algae and plants, substantial variation is also found in the sizes of chloroplast chromosomes. A typical chloroplast genome is approximately 100,000 to 200,000 bp in length. **Figure 6.2** shows the **chloroplast DNA (cpDNA)** of the tobacco plant, which is a circular DNA molecule that contains 156,000 bp of DNA and carries between 110 and 120 different genes. By comparison, the nuclear genome of tobacco contains tens of thousands of genes. Some chloroplast genes encode ribosomal RNAs and transfer RNAs, which are required for protein synthesis inside the chloroplast. Other chloroplast genes encode proteins involved with photosynthesis. These proteins are made inside the chloroplast. In addition, many chloroplast proteins are encoded by genes found in the plant cell nucleus. These nuclear-encoded proteins are made in the cytosol and contain chloroplast-targeting signals that direct them into the chloroplasts.

Extranuclear Inheritance Produces Non-Mendelian Results in Reciprocal Crosses

In diploid eukaryotic species, most genes within the nucleus obey a Mendelian pattern of inheritance because the homologous pairs of chromosomes segregate during meiosis. Except for sex-linked traits, offspring inherit one copy of each gene from both the maternal and paternal parents. The sorting of chromosomes during meiosis explains the inheritance patterns of **nuclear genes,** which are genes on chromosomes in the cell nucleus. By comparison, the inheritance of extranuclear genetic material does not display a Mendelian pattern. Chloroplasts and mitochondria are not sorted during meiosis and therefore do not segregate into gametes in the same way as nuclear chromosomes.

In 1909, Carl Correns discovered a trait that showed a non-Mendelian pattern of inheritance involving pigmentation in *Mirabilis jalapa* (the four-o'clock plant). Leaves can be green, white, or variegated with both green and white sectors. Correns demonstrated that the pigmentation of the offspring depended solely on the maternal parent (**Figure 6.3**). If the female parent had white pigmentation, all offspring had white leaves. Similarly, if the female was green, all offspring were green. When the female was variegated, the offspring could be green, white, or variegated.

The pattern of inheritance observed by Correns is a type of extranuclear inheritance called **maternal inheritance.** Chloroplasts are a type of plastid that makes chlorophyll, a green photosynthetic pigment. Maternal inheritance occurs because the chloroplasts are inherited only through the cytoplasm of the egg. The sperm cells within pollen grains of *M. jalapa* do not transmit chloroplasts to the offspring.

The phenotypes of leaves can be explained by the types of chloroplasts within the leaf cells. The green phenotype, which is the wild-type condition, is due to the presence of normal chloroplasts that make green pigment. By comparison, the white phenotype is due to a mutation in a gene within the chloroplast DNA that diminishes the synthesis of green pigment. A cell may contain both types of chloroplasts, a condition known as

TABLE 6.1

Genetic Composition of Chloroplasts

Species	Organelle	Nucleoids per Organelle	Total Number of Chromosomes per Organelle
Chlamydomonas	Chloroplast	5–6	~80
Euglena	Chloroplast	20–34	100–300
Flowering plants	Chloroplast	12–25	~60

Data from: N.M. Gillham (1994). *Organelle Genes and Genomes.* Oxford University Press, New York.

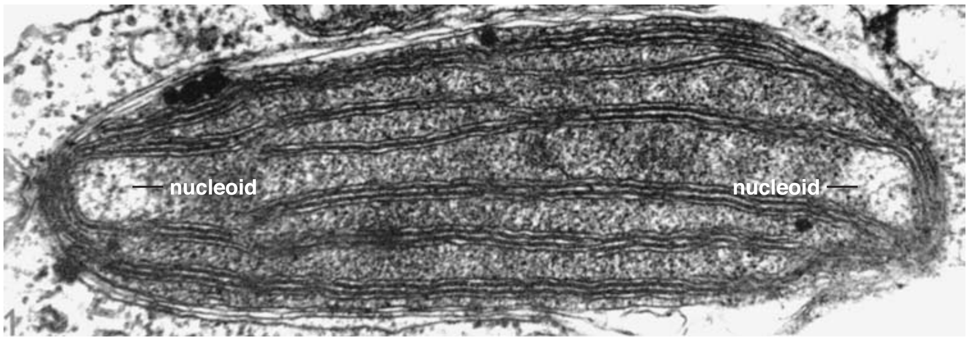


FIGURE 6.1 **Nucleoids within a chloroplast.** The chloroplast chromosomes are found within the nucleoid region of the organelle.

Concept check: How is a nucleoid different from a cell nucleus?

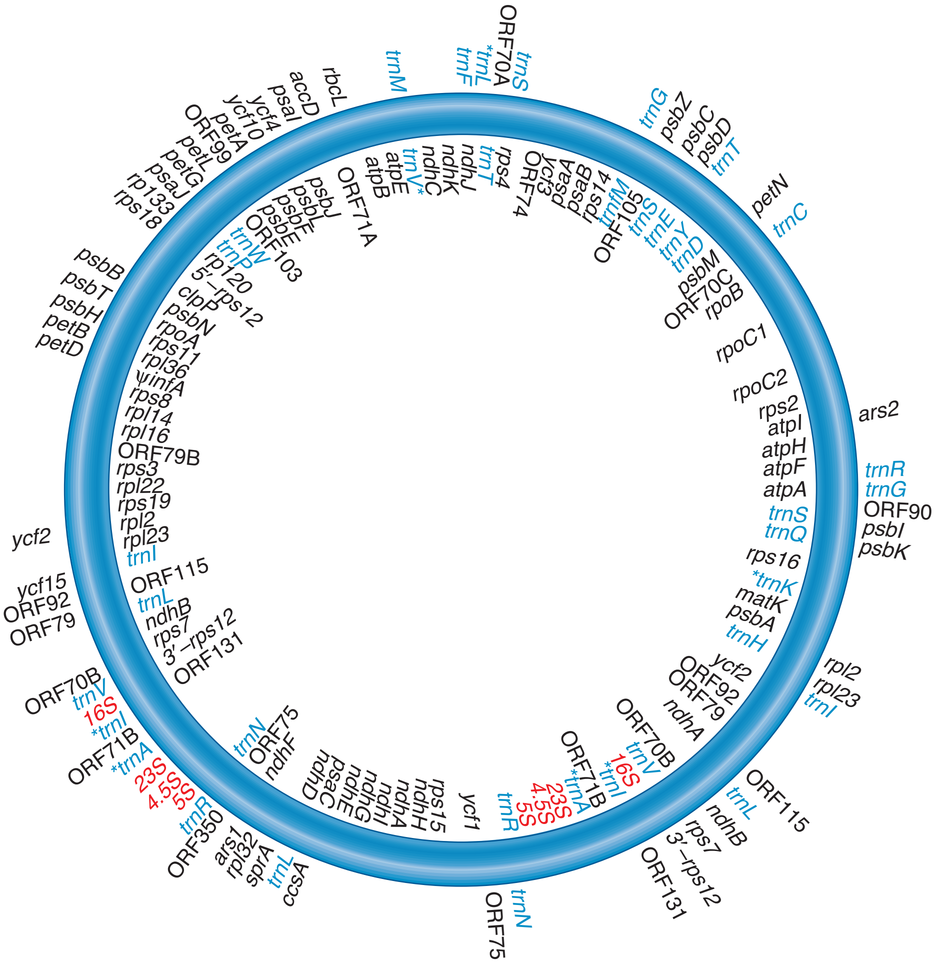
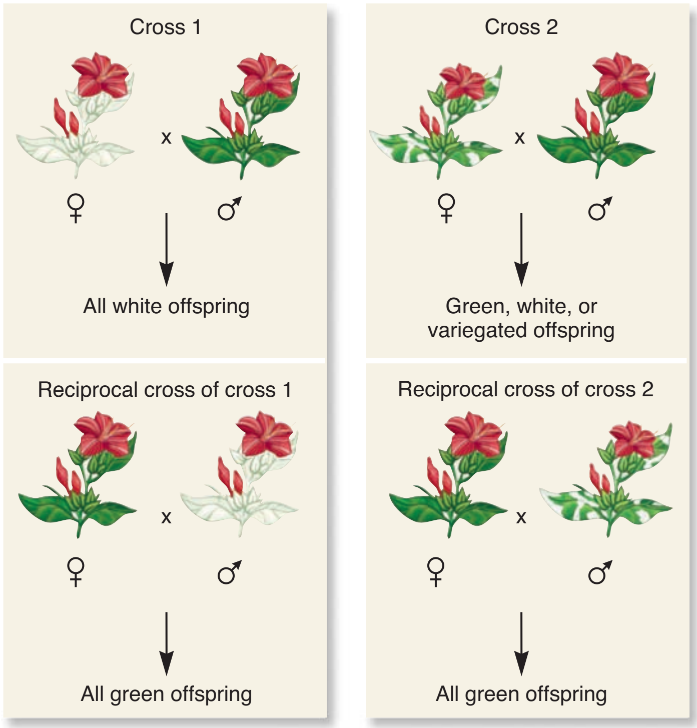

trnS
ORF70A
*trnL
trnF
trnM
rbcL
accD
psaI
ycf4
ycf10
petA
ORF99
petL
petG
psaJ
rpl33
rps18
psbB
psbT
psbH
petB
petD
ycf2
ycf15
ORF92
ORF79
ORF70B
trnV
16S
*trnI
ORF71B
*trnA
23S
4.5S
5S
trnR
ORF350
ars1
rpl32
sprA
trnL
ccsA
ORF75
trnN
ORF131
3′–rps12
rps7
ndhB
trnL
ORF115
trnI
rpl23
rpl2
psbK
psbI
ORF90
trnG
trnR
ars2
trnC
petN
trnT
psbD
psbC
psbZ
trnG
ORF71A
atpB
atpE
trnV*
ndhC
ndhK
ndhJ
trnT
rps4
ORF74
ycf3
psaA
psaB
rps14
trnfM
ORF105
trnS
trnE
trnY
trnD
psbM
ORF70C
rpoB
rpoC1
rpoC2
rps2
atpI
atpH
atpF
atpA
trnS
trnQ
rps16
*trnK
matK
psbA
trnH
ycf2
ORF92
ORF79
ndhA
ORF70B
trnV
16S
*trnI
ORF71B
*trnA
23S
4.5S
5S
trnR
ycf1
rps15
ndhH
ndhA
ndhI
ndhG
ndhE
psaC
ndhD
ndhF
ORF75
trnN
ORF131
3′–rps12
rps7
ndhB
trnL
ORF115
trnI
rpl23
rpl2
rps19
rpl22
rps3
ORF79B
rpl16
rpl14
rps8
ψinfA
rpl36
rps11
rpoA
psbN
clpP
5′–rps12
rpl20
trnP
trnW
ORF103
psbE
psbF
psbL
psbJ

FIGURE 6.2 **A genetic map of tobacco chloroplast DNA (cpDNA).** This diagram illustrates the locations of many genes along the circular chloroplast chromosome. The gene names shown in blue encode transfer RNAs. The genes that encode ribosomal RNA are shown in red. The remaining genes shown in black encode polypeptides that function within the chloroplast. The genes designated ORF (open reading frame) encode polypeptides with unknown functions.

Concept check: With regard to the total number of genes, how does the chloroplast genome compare to the nuclear genome?

heteroplasmy. A leaf cell containing both types of chloroplasts is green because the normal chloroplasts produce green pigment.

How does a variegated phenotype occur? **Figure 6.4** considers the leaf of a plant that began from a fertilized egg that contained both types of chloroplasts (i.e., a heteroplasmic cell). As a plant grows, the two types of chloroplasts are irregularly distributed to daughter cells. On occasion, a cell may receive only the chloroplasts that are defective in making green pigment. Such a cell continues to divide and produces a sector of the plant that is entirely white. In this way, the variegated phenotype is produced.


Cross 1
x
♀
♂
All white offspring
Cross 2
x
♀
♂
Green, white, or variegated offspring
Reciprocal cross of cross 1
x
♀
♂
All green offspring
Reciprocal cross of cross 2
x
♀
♂
All green offspring

INTERACTIVE EXERCISE

FIGURE 6.3 **Maternal inheritance in the four-o'clock plant, *Mirabilis jalapa*.** The reciprocal crosses of four-o'clock plants by Carl Correns consisted of a pair of crosses between white-leaved and green-leaved plants, and a pair of crosses between variegated-leaved and green-leaved plants.

Genes → Traits In this example, the white phenotype is due to chloroplasts that carry a mutant allele that diminishes green pigmentation. The variegated phenotype is due to a mixture of chloroplasts, some of which carry the normal (green) allele and some of which carry the white allele. In the crosses shown here, the parent providing the eggs determines the phenotypes of the offspring. This is due to maternal inheritance. The egg contains the chloroplasts that are inherited by the offspring. (Note: The defective chloroplasts that give rise to white sectors are not completely defective in chlorophyll synthesis. Therefore, entirely white plants can survive, though they are smaller than green or variegated plants.)

Concept check: What is a reciprocal cross?

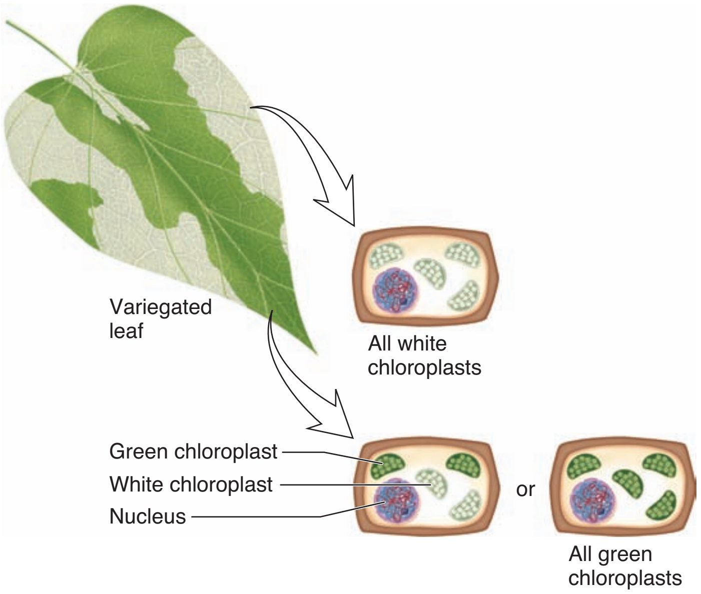


FIGURE 6.4 **A cellular explanation of the variegated phenotype in *Mirabilis jalapa*.** This plant inherited two types of chloroplasts—those that can produce green pigment and those that are defective. As the plant grows, the two types of chloroplasts are irregularly distributed to daughter cells. On occasion, a leaf cell may receive only the chloroplasts that are defective at making green pigment. Such a cell continues to divide and produces a sector of the leaf that is entirely white. Cells that contain both types of chloroplasts or cells that contain only green chloroplasts produce green tissue, which may be adjacent to a sector of white tissue. This is the basis for the variegated phenotype of the leaves.

Concept check: During growth, can a patch of tissue with a white phenotype give rise to a patch with a green phenotype? Explain.

Similarly, if we consider the results of Figure 6.3, a female parent that is variegated may transmit green, white, or a mixture of these types of chloroplasts to the egg cell, thereby producing green, white, or variegated offspring, respectively.

Studies in *Chlamydomonas* Provided Genetic Evidence for Extranuclear Inheritance of Chloroplasts

The research of Correns and others indicated that some traits, such as leaf pigmentation, are inherited in a non-Mendelian manner. However, such studies did not definitively determine that maternal inheritance is due to genetic material within organelles. Further progress in the investigation of extranuclear inheritance was provided by detailed genetic analyses of eukaryotic microorganisms, such as algae, by isolating and characterizing mutant phenotypes that specifically affected the chloroplasts.

The unicellular alga *Chlamydomonas reinhardtii* has been used as a model organism to investigate the inheritance of chloroplasts. This organism contains a single chloroplast that occupies approximately 40% of the cell volume. Genetic studies of chloroplast inheritance began when Ruth Sager identified a mutant strain of *Chlamydomonas* that is resistant to the antibiotic streptomycin (sm^r). By comparison, most strains are sensitive to killing by streptomycin (sm^s).

Sager conducted crosses to determine the inheritance pattern of the sm^r gene. During mating, two haploid cells unite to form a diploid cell, which then undergoes meiosis to form four haploid cells. *Chlamydomonas* is an organism that can be found in two mating types, in this case, designated mt^+ and mt^-. Mating type is due to nuclear inheritance and segregates in a 1:1 manner. By comparison, Sager and her colleagues discovered that the sm^r gene was inherited from the mt^+ parent but not from the mt^- parent (**Figure 6.5**). Therefore, this sm^r gene was not inherited in a Mendelian manner. Further research revealed that this pattern occurred because only the mt^+ parent transmits chloroplasts to daughter cells and the sm^r gene is found in the chloroplast genome.

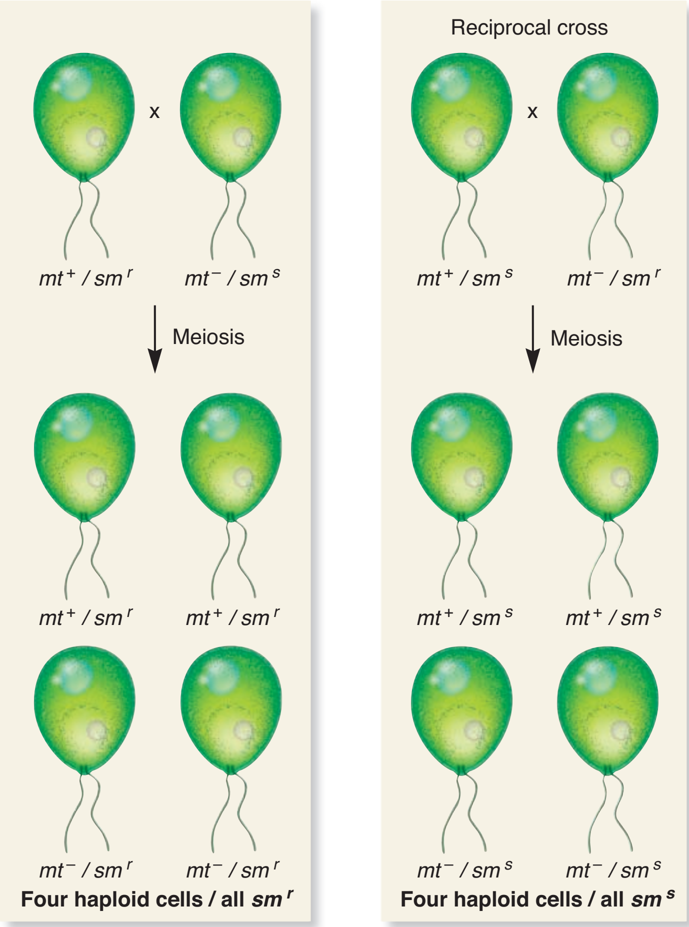


FIGURE 6.5 **Chloroplast inheritance in *Chlamydomonas*.** The two mating types of the organism are designated mt^+ and mt^-. Sm^r indicates streptomycin resistance, whereas sm^s indicates sensitivity to this antibiotic.

Concept check: In this species, which mating type transmits chloroplasts to offspring?

TABLE 6.2

Transmission of Chloroplasts Among Different Species

Species	Organelle	Transmission
Chlamydomonas	Chloroplasts	Inherited from the parent with the *mt*$^+$ mating type
Plants		
Angiosperms	Chloroplasts	Often maternal inheritance, although biparental inheritance* is found among some species
Gymnosperms	Chloroplasts	Usually paternal inheritance**

*A pattern in which the organelle is transmitted from both parents.

**A pattern in which the organelle is transmitted only via sperm cells within pollen grains.

The Pattern of Inheritance of Chloroplasts Varies Among Different Species

The inheritance of traits via genetic material within chloroplasts is now a well-established phenomenon that geneticists have investigated in many different species. In **heterogamous** species, two morphologically different types of gametes are made—eggs and sperm. Egg cells tend to be large and provide most of the cytoplasm to the zygotes, whereas the sperm are small and often provide little more than a nucleus. Therefore, chloroplasts are most often inherited from the maternal parent. However, this is not always the case. **Table 6.2** describes the inheritance patterns of chloroplasts in selected species.

6.1 REVIEWING THE KEY CONCEPTS

- Extranuclear inheritance involves the inheritance of genes that are found in mitochondria or chloroplasts.
- Chloroplasts carry circular chromosomes in a nucleoid region. These circular chromosomes contain many genes but far fewer compared with the number in the cell nucleus (see Figures 6.1, 6.2, Table 6.1).
- Maternal inheritance occurs when organelles, such as chloroplasts, are transmitted via the egg (see Figure 6.3).
- Heteroplasmy for chloroplasts can result in a variegated phenotype (see Figure 6.4).
- In the alga *Chlamydomonas*, chloroplasts are transmitted from the *mt*$^+$ parent (see Figure 6.5).
- The transmission patterns of chloroplasts vary among different species (see Table 6.2).

6.1 COMPREHENSION QUESTIONS

1. Extranuclear inheritance occurs due to
   a. chromosomes that may become detached from the spindle during meiosis.
   b. genetic material that is found in chloroplasts and mitochondria.
   c. mutations that disrupt the integrity of the nuclear membrane.
   d. none of the above.
2. The genetic material of a chloroplast is found in the ____________ and consists of __________.
   a. nucleus, several genes
   b. nucleus, one or more copies of a circular chromosome
   c. nucleoid, a few genes
   d. nucleoid, one or more copies of a circular chromosome
3. A cross is made between a green four-o'clock and a variegated four o'clock. If the variegated plant provides the pollen, the expected outcome of the offspring would be
   a. all plants with green leaves.
   b. 3 plants with green leaves to 1 plant with variegated leaves.
   c. 3 plants with green leaves to 1 plant with white leaves.
   d. some plants with green leaves, some with variegated leaves, and some with white leaves.

6.2 EXTRANUCLEAR INHERITANCE: MITOCHONDRIA

As we have seen, chloroplasts contain their own genetic material and their transmission pattern is different from a traditional Mendelian pattern. A similar situation exists for mitochondria. In this section, we will examine the genetic composition of mitochondria and explore the pattern of transmission of this organelle from parent to offspring.

Mitochondria Also Contain Circular Chromosomes with Many Genes

Like chloroplasts, the genetic material of mitochondria is located inside the organelle in a region known as the nucleoid (**Figure 6.6**). The genome is a single circular chromosome, although a nucleoid usually contains multiple copies of this chromosome. In addition, a mitochondrion often has more than one nucleoid. In mice, for example, each mitochondrion has one to three nucleoids, with each nucleoid containing two to six copies of the circular mitochondrial genome. However, this number varies depending on the type of cell and the stage of development.

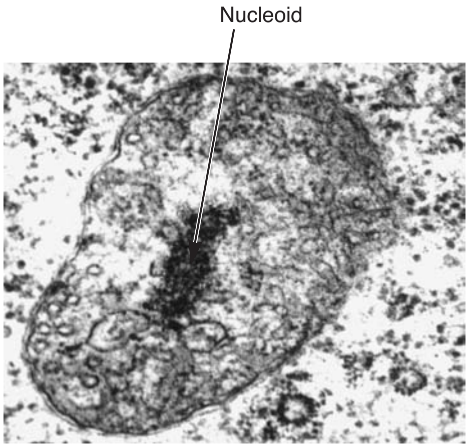


FIGURE 6.6 **Nucleoid within a mitochondrion.** The mitochondrial chromosomes are found within the nucleoid region of the organelle.

Concept check: Within one nucleoid, is the mitochondrial chromosome found in a single copy or multiple copies?

TABLE **6.3**

Genetic Composition of Mitochondria

Species	Organelle	Nucleoids per Organelle	Total Number of Chromosomes per Organelle
Tetrahymena	Mitochondrion	1	6–8
Mouse	Mitochondrion	1–3	5–6

Data from: N.W. Gillham (1994). *Organelle Genes and Genomes.* Oxford University Press, New York.

Table 6.3 describes the genetic composition of mitochondria for two selected species.

Besides variation in copy number, the sizes of mitochondrial genomes also vary greatly among different species. A 400-fold variation is found in the sizes of mitochondrial chromosomes. In general, the mitochondrial genomes of animal species tend to be fairly small; those of fungi and protists are intermediate in size; and those of plant cells tend to be fairly large.

Figure 6.7 illustrates a map of human **mitochondrial DNA (mtDNA).** Each copy of the mitochondrial chromosome consists of a circular DNA molecule that is only 17,000 bp in length. This size is less than 1% of a typical bacterial chromosome. The human mtDNA carries relatively few genes. Thirteen genes encode polypeptides that function within the mitochondrion. In addition, mtDNA carries genes that encode ribosomal RNA and transfer RNA, which are needed for the synthesis of the 13 polypeptides that are encoded by the mtDNA. The primary role of mitochondria is to provide cells with the bulk of their adenosine triphosphate (ATP), which is used as an energy source to drive cellular reactions. These 13 polypeptides are subunits of proteins that function in a process known as oxidative phosphorylation, in which mitochondria use oxygen and synthesize ATP. However, mitochondria require many additional proteins to carry out oxidative phosphorylation and other mitochondrial functions. Most mitochondrial proteins are encoded by genes within the cell nucleus and contain mitochondrial-targeting signals.

TABLE **6.4**

Transmission of Mitochondria Among Different Species

Species	Organelle	Transmission
Mammals	Mitochondria	Maternal inheritance
S. cerevisiae	Mitochondria	Biparental inheritance
Molds	Mitochondria	Usually maternal inheritance; paternal inheritance has been found in the genus *Allomyces*
Chlamydomonas	Mitochondria	Inherited from the parent with the *mt*⁻ mating type
Plants		
Angiosperms	Mitochondria	Often maternal inheritance, although biparental inheritance is found among some species
Gymnosperms	Mitochondria	Usually paternal inheritance

The Transmission of Mitochondria Usually Follows a Maternal Inheritance Pattern

In heterogamous species that make female and male gametes, mitochondria are usually inherited via egg cells. Therefore, the female parent passes mitochondrial genes to her offspring, but the male parent does not. However, this is not always the case. **Table 6.4** describes the inheritance patterns of mitochondria in selected species.

In species in which maternal inheritance is generally observed, the paternal parent may occasionally provide mitochondria via the sperm. This phenomenon, called **paternal leakage,** occurs in many species that primarily exhibit maternal inheritance

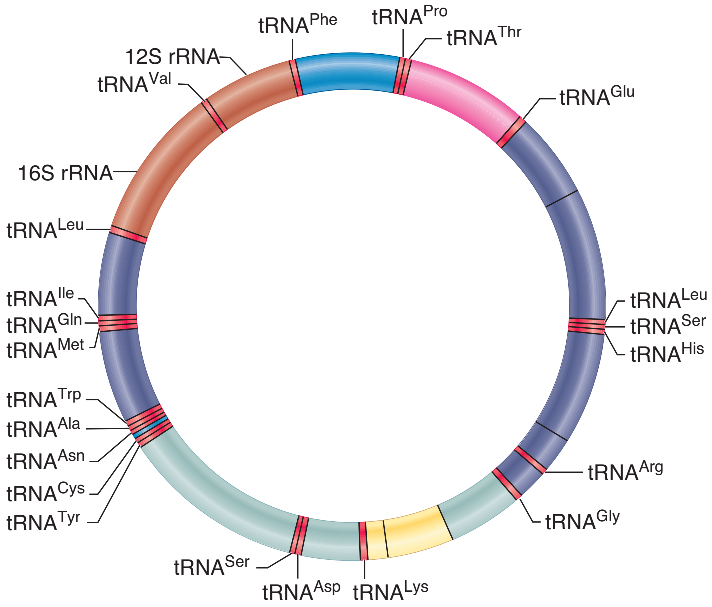


FIGURE 6.7 **A genetic map of human mitochondrial DNA (mtDNA).** This diagram illustrates the locations of many genes along the circular mitochondrial chromosome. The genes shown in red encode transfer RNAs. For example, tRNA^{Arg} encodes a tRNA that carries arginine. The genes that encode ribosomal RNA are shown in light brown. The remaining genes encode proteins that function within the mitochondrion. The mitochondrial genomes from numerous species have been determined.

Concept check: *Why do mitochondria need rRNA and tRNA genes?*

of their organelles. In the mouse, for example, approximately 1 to 4 paternal mitochondria are inherited for every 100,000 maternal mitochondria per generation of offspring. Most offspring do not inherit any paternal mitochondria, but a rare individual may inherit a mitochondrion from the sperm.

Many Human Diseases Are Caused by Mitochondrial Mutations

Researchers have identified several human diseases that are caused by mutations in mtDNA. These diseases can occur in two ways. In some cases, mitochondrial mutations that cause disease are transmitted from mother to offspring. Human mtDNA is maternally inherited because it is transmitted from mother to offspring via the cytoplasm of the egg. Therefore, the transmission of inherited human mitochondrial diseases follows a maternal inheritance pattern. In addition, mitochondrial mutations may occur in somatic cells and accumulate as a person ages. Researchers have discovered that mitochondria are particularly susceptible to DNA damage. When more oxygen is consumed than is actually used to make ATP, mitochondria tend to produce free radicals that damage DNA. Unlike nuclear DNA, mitochondrial DNA has very limited repair abilities and almost no protective ability against free-radical damage.

Table 6.5 describes several mitochondrial diseases that have been discovered in humans and are caused by mutations in mitochondrial genes. Over 200 diseases associated with defective mitochondria have been discovered. These are usually chronic degenerative disorders that affect cells requiring a high level of ATP, such as nerve and muscle cells. For example, Leber hereditary optic neuropathy (LHON) affects the optic nerve and may lead to the progressive loss of vision in one or both eyes. LHON can be caused by a defective mutation in one of several different mitochondrial genes. Researchers are still investigating how a defect in these mitochondrial genes produces the symptoms of this disease.

An important factor in mitochondrial disease is heteroplasmy, which means that a cell contains a mixed population of mitochondria. Within a single cell, some mitochondria may carry a disease-causing mutation, whereas others may not. As cells divide, mutant and normal mitochondria randomly segregate into the resulting daughter cells. Some daughter cells may receive a high ratio of mutant to normal mitochondria, whereas others may receive a low ratio. The ratio of mutant to normal mitochondria must exceed a certain threshold value before disease symptoms are observed.

TABLE **6.5**

Examples of Human Mitochondrial Diseases

Disease	Mitochondrial Gene Mutated
Leber hereditary optic neuropathy	A mutation in one of several mitochondrial genes that encode respiratory chain proteins: *ND1, ND2, CO1, ND4, ND5, ND6,* and *cytb*
Neurogenic muscle weakness	A mutation in the *ATPase6* gene that encodes a subunit of the mitochondrial ATP-synthetase, which is required for ATP synthesis
Mitochondrial myopathy	A mutation in a gene that encodes a tRNA for leucine
Maternal myopathy and cardiomyopathy	A mutation in a gene that encodes a tRNA for leucine
Myoclonic epilepsy with ragged-red muscle fibers	A mutation in a gene that encodes a tRNA for lysine

6.2 REVIEWING THE KEY CONCEPTS

- Mitochondria carry circular chromosomes in a nucleoid region (see Figures 6.6, 6.7, Table 6.3).
- Mitochondria are usually transmitted by a maternal inheritance pattern, but transmission patterns vary among different species (see Table 6.4).
- Many human diseases are caused by mutations in mitochondrial DNA (see Table 6.5).

6.2 COMPREHENSION QUESTIONS

1. Human mitochondrial DNA
    a. is found in the nucleoid region of the mitochondrion.
    b. is found in a circular chromosome.
    c. encodes rRNAs, tRNA, and several proteins.
    d. is all of the above.
2. In most species, the transmission of mitochondria follows
    a. a maternal inheritance pattern.
    b. a paternal inheritance pattern.
    c. a biparental inheritance pattern.
    d. none of the above.
3. Some human diseases are caused by mutations in mitochondrial genes. Which of the following statements is *false*?
    a. Mitochondrial diseases may follow a maternal inheritance pattern.
    b. Mutations associated with mitochondrial diseases often affect cells with a high demand for ATP.
    c. The symptoms associated with mitochondrial diseases tend to improve with age.
    d. Heteroplasmy plays a key role in the severity of disease symptoms.

6.3 THEORY OF ENDOSYMBIOSIS

The idea that the nucleus, chloroplasts, and mitochondria contain their own separate genetic material may at first seem puzzling. Wouldn't it be simpler to have all of the genetic material in one place in the cell? The underlying reason for distinct genomes of chloroplasts and mitochondria can be traced back to their evolutionary origin, which is thought to involve a symbiotic association.

A symbiotic relationship occurs when two different species live together in a close association. The symbiont is the smaller of the two species; the host is the larger. The term **endosymbiosis** describes a symbiotic relationship in which the symbiont actually lives inside (*endo*-, inside) the host. In 1883, Andreas Schimper proposed that chloroplasts were descended from an endosymbiotic relationship between cyanobacteria and eukaryotic cells. This idea, now known as the **endosymbiosis theory,** suggested that the ancient origin of chloroplasts was initiated when a cyanobacterium took up residence within a primordial eukaryotic cell (**Figure 6.8**). Over the course of evolution, the characteristics of the intracellular bacterial cell gradually changed to those of a chloroplast. In 1922, Ivan Wallin also proposed an endosymbiotic origin for mitochondria.

In spite of these hypotheses, the question of endosymbiosis was largely ignored until researchers in the 1950s discovered that chloroplasts and mitochondria contain their own genetic material. The issue of endosymbiosis was hotly debated after Lynn Margulis published a book entitled *Origin of Eukaryotic Cells* (1970), in which she proposed that chloroplasts and mitochondria evolved from endosymbiotic bacteria. During the 1970s and 1980s, the advent of molecular genetic techniques allowed researchers to analyze genes from chloroplasts, mitochondria, bacteria, and eukaryotic nuclear genomes. They found that genes in chloroplasts and mitochondria are very similar to bacterial genes but not as similar to those found within the nucleus of eukaryotic cells. This observation provided strong support for the endosymbiotic origin of chloroplasts and mitochondria, which is now widely accepted.

The endosymbiosis theory proposes that the relationship provided eukaryotic cells with useful cellular characteristics. Chloroplasts were derived from cyanobacteria, a bacterial species that is capable of photosynthesis. The ability to carry out photosynthesis enabled algal and plant cells to use the energy from sunlight. By comparison, mitochondria are thought to have been derived from a different type of bacteria known as gram-negative nonsulfur purple bacteria. In this case, the endosymbiotic relationship enabled eukaryotic cells to synthesize greater amounts of ATP. It is less clear how the relationship would have been beneficial to cyanobacteria or purple bacteria, though the cytosol of a eukaryotic cell may have provided a stable environment with an adequate supply of nutrients.

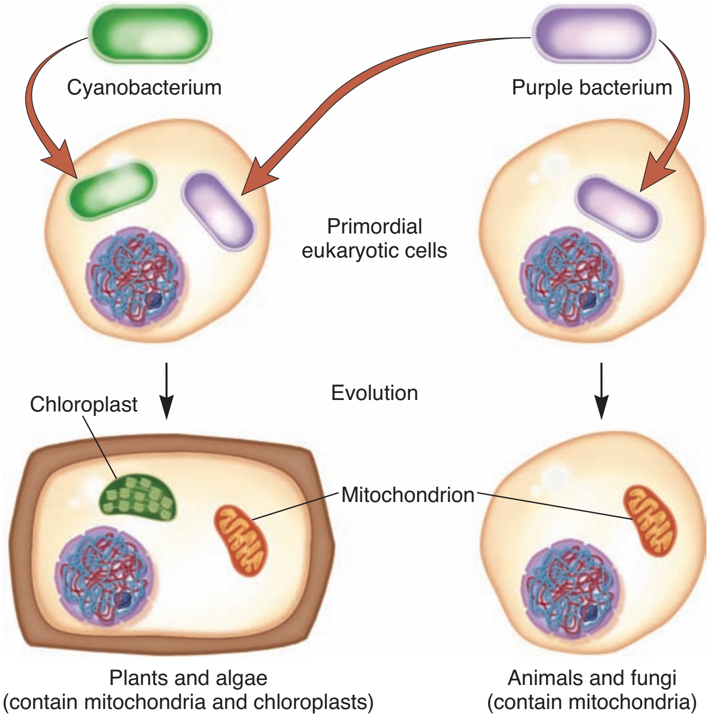

Cyanobacterium
Purple bacterium
Primordial eukaryotic cells
Chloroplast
Evolution
Mitochondrion
Plants and algae (contain mitochondria and chloroplasts)
Animals and fungi (contain mitochondria)

FIGURE 6.8 The endosymbiotic origin of chloroplasts and mitochondria. According to the endosymbiotic theory, chloroplasts descended from an endosymbiotic relationship between cyanobacteria and eukaryotic cells. This arose when a cyanobacterium took up residence within a primordial eukaryotic cell. Over the course of evolution, the intracellular bacterial cell gradually changed its characteristics, eventually becoming a chloroplast. Similarly, mitochondria are derived from an endosymbiotic relationship between purple bacteria and eukaryotic cells.

Concept check: How have chloroplasts and mitochondria changed since the initial endosymbiosis events, which occurred hundreds of millions of years ago?

During the evolution of eukaryotic species, most genes that were originally found in the genome of the primordial cyanobacteria and purple bacteria have been lost or transferred from the organelles to the nucleus. The sequences of certain genes within the nucleus are consistent with their origin within an organelle. Such genes are more similar in their DNA sequence to known bacterial genes than to their eukaryotic counterparts. Therefore, researchers have concluded that these genes have been removed from the chloroplast and mitochondrial chromosomes and relocated to the nuclear chromosomes. This has occurred many times throughout evolution, so modern chloroplasts and mitochondria have lost most of the genes that are still found in present-day cyanobacteria and purple bacteria.

Most of this gene transfer occurred early in chloroplast and mitochondrial evolution. The functional transfer of mitochondrial genes seems to have ceased in animals, but gene transfer from chloroplasts and mitochondria to the nucleus continues to occur in plants at a low rate. The molecular mechanism of gene transfer is not entirely understood, but the direction of transfer is well established. During evolution, gene transfer has occurred primarily from the organelles to the nucleus. For example, about 1500 genes have been transferred from the mitochondrial genome to the nuclear genome of eukaryotes. Transfer of genes from the nucleus to the organelles has almost never occurred, although one example is known of a nuclear gene in plants that has been transferred to the mitochondrial genome. This unidirectional gene transfer from organelles to the nucleus partly explains why the organellar genomes now contain relatively few genes. In addition, gene transfer can occur between organelles. It can happen between two mitochondria, between two chloroplasts, and between a chloroplast and mitochondrion. Overall, the transfer of genetic material between the nucleus, chloroplasts, and mitochondria is an established phenomenon, although its biological benefits remain unclear.

6.3 REVIEWING THE KEY CONCEPTS

- Chloroplasts and mitochondria were derived from an ancient endosymbiotic relationship (see Figure 6.8).

6.3 COMPREHENSION QUESTIONS

1. Chloroplasts and mitochondria evolved from an endosymbiotic relationship involving
   a. purple bacteria and cyanobacteria, respectively.
   b. cyanobacteria and purple bacteria, respectively.
   c. cyanobacteria.
   d. purple bacteria.

6.4 EPIGENETICS: IMPRINTING

We now turn to a non-Mendelian pattern of inheritance that involves nuclear genes. **Epigenetic inheritance** is a pattern in which a modification occurs to a nuclear gene or chromosome that alters gene expression, but is not permanent over the course of many generations. Epigenetic inheritance patterns are the result of DNA and chromosomal modifications that occur during oogenesis, spermatogenesis, or early stages of embryogenesis. Once they are initiated during these early stages, epigenetic changes alter the expression of particular genes in a way that may be fixed during an individual's lifetime. Therefore, epigenetic changes can permanently affect the phenotype of the individual. However, epigenetic modifications are not permanent over the course of many generations, and they do not change the actual DNA sequence. For example, a gene may undergo an epigenetic change that inactivates it for the lifetime of an individual. However, when this individual makes gametes, the gene may become activated and remain operative during the lifetime of an offspring who inherits the active gene.

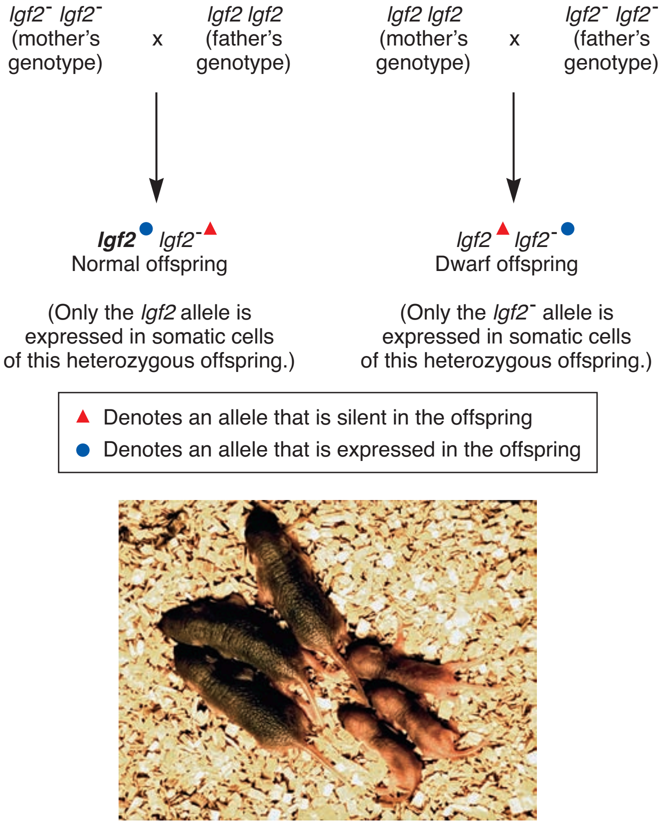
We already considered one example of epigenetic inheritance, X inactivation, in Chapter 4 (see Figures 4.5–4.7). During this process, one of the two X chromosomes in female mammals is chosen for random inactivation during embryonic development. This may lead to a variegated phenotype in females. In this section, we will examine another form of epigenetic inheritance called **genomic imprinting,** or simply **imprinting,** in which a modification occurs to a nuclear gene that alters gene expression, but is not permanent over the course of many generations.

The Expression of an Imprinted Gene Depends on the Sex of the Parent from Which the Gene Was Inherited

The term *imprinting* implies a type of marking process that has a memory. For example, newly hatched birds identify marks on their parents that allow them to distinguish their parents from other individuals. The term *genomic imprinting* refers to an analogous situation in which a segment of DNA is marked, and that mark is retained and recognized throughout the life of the organism inheriting the marked DNA. The phenotypes caused by imprinted genes follow a non-Mendelian pattern of inheritance because the marking process causes the offspring to distinguish between maternally and paternally inherited alleles. Depending on how the genes are marked, the offspring expresses only one of the two alleles. This phenomenon is termed **monoallelic expression.**

To understand genomic imprinting, let's consider a specific example. In the mouse, a gene designated *Igf2* encodes a protein growth hormone called insulin-like growth factor 2. Imprinting results in the expression of the paternal *Igf2* allele but not the maternal allele. The paternal allele is transcribed into RNA, but the maternal allele is transcriptionally silent. With regard to phenotype, a functional *Igf2* gene is necessary for normal size. A loss-of-function allele of this gene, designated *Igf2*$^{-}$, is defective in the synthesis of a functional Igf2 protein. This may cause a mouse to be a dwarf, but the dwarfism depends on whether the mutant allele is inherited from the male or female parent, as shown in

Figure 6.9. On the left side, an offspring has inherited the *Igf2* allele from its father and the *Igf2*$^{-}$ allele from its mother. Due to imprinting, only the *Igf2* allele is expressed in the offspring. Therefore, this mouse grows to a normal size. Alternatively, in the reciprocal cross on the right side, an individual has inherited the *Igf2*$^{-}$ allele from its father and the *Igf2* allele from its mother. In this case, the *Igf2* allele is not expressed. In this mouse, the *Igf2*$^{-}$ allele is transcribed into mRNA, but the mutation renders the Igf2 protein defective. Therefore, the offspring on the right has a dwarf phenotype. As shown here, both offspring have the same genotype; they are heterozygous for the *Igf2* alleles (i.e., *Igf2*


Igf2⁻ Igf2⁻ (mother's genotype)
x
Igf2 Igf2 (father's genotype)
Igf2 Igf2 (mother's genotype)
x
Igf2⁻ Igf2⁻ (father's genotype)
Igf2 ● Igf2⁻ ▲
Normal offspring
(Only the Igf2 allele is expressed in somatic cells of this heterozygous offspring.)
Igf2 ▲ Igf2⁻ ●
Dwarf offspring
(Only the Igf2⁻ allele is expressed in somatic cells of this heterozygous offspring.)
▲ Denotes an allele that is silent in the offspring
● Denotes an allele that is expressed in the offspring

INTERACTIVE EXERCISE

FIGURE 6.9 An example of genomic imprinting in the mouse. In the cross on the left, a homozygous male with the normal *Igf2* allele is crossed to a homozygous female carrying a defective allele, designated *Igf2*$^{-}$. An offspring is heterozygous and normal because the paternal allele is active. In the reciprocal cross on the right, a homozygous male carrying the defective allele is crossed to a homozygous normal female. In this case, the offspring is heterozygous and dwarf. This is because the paternal allele is defective due to a mutation and the maternal allele is not expressed. The photograph shows normal-size (left) and dwarf littermates (right) derived from a cross between a wild-type female and a heterozygous male carrying a loss-of-function *Igf2* allele (courtesy of A. Efstratiadis). The loss-of-function allele was created using gene knockout methods described in Chapter 20 (see Figure 20.3a).

Concept check: What would be the outcome of a cross between a heterozygous female and a male that carries two normal copies of the Igf2 gene?

Igf2$^-$). They are phenotypically different, however, because only the paternally inherited allele is expressed.

At the cellular level, imprinting is a process that can be divided into three stages: (1) the establishment of the imprint during gametogenesis, (2) the maintenance of the imprint during embryogenesis and in adult somatic cells, and (3) the erasure and reestablishment of the imprint in the germ cells. These stages are described in **Figure 6.10**, which shows the imprinting of the *Igf2* gene. The two mice shown here have inherited the *Igf2* allele from their father and the *Igf2*$^-$ allele from their mother. Due to imprinting, both mice express the *Igf2* allele in their somatic cells, and the pattern of imprinting is maintained in the somatic cells throughout development. In the germ cells (i.e., sperm and eggs), the imprint is erased; it is reestablished according to the sex of the animal. The female mouse on the left transmits only transcriptionally inactive alleles to her offspring. The male mouse on the right transmits transcriptionally active alleles. However, because this male is a heterozygote, it transmits either a functionally active *Igf2* allele or a functionally defective mutant allele (*Igf2*$^-$). An *Igf2*$^-$ allele, which is inherited from a male mouse, can be expressed into mRNA (i.e., it is transcriptionally active), but it does not produce a functional Igf2 protein due to the deleterious mutation that created the *Igf2*$^-$ allele; a dwarf phenotype results.

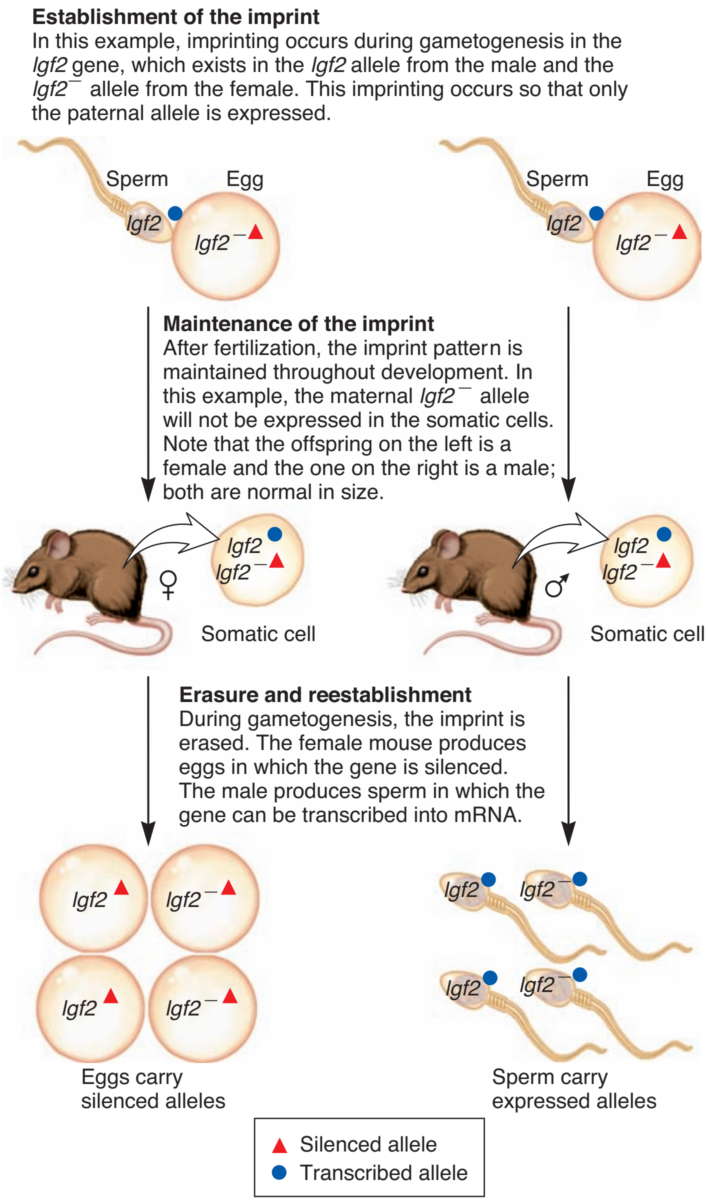
As seen in Figure 6.10, genomic imprinting is permanent in the somatic cells of an animal, but the marking of alleles can be altered from generation to generation. For example, the female mouse on the left has an active copy of the *Igf2* allele, but any allele this female transmits to its offspring is transcriptionally inactive.

Genomic imprinting occurs in several species, including numerous insects, mammals, and flowering plants. Imprinting may involve a single gene, a part of a chromosome, an entire chromosome, or even all of the chromosomes from one parent. Helen Crouse discovered the first example of imprinting, which involved an entire chromosome in the housefly, *Sciara coprophila*. In this species, the fly normally inherits three sex chromosomes, rather than two as in most other species. One X chromosome is inherited from the female, and two are inherited from the male. In male flies, both paternal X chromosomes are lost from somatic cells during embryogenesis. In female flies, only one of the paternal X chromosomes is lost. In both sexes, the maternally inherited X chromosome is never lost. These results indicate that either the maternal X chromosome is marked to promote its retention or the paternal X chromosomes are marked to promote their loss.

Genomic imprinting can also be involved in the process of X inactivation, described previously in Chapter 4. In certain species, imprinting plays a role in the choice of which X chromosome will be inactivated. For example, in marsupials, the paternal X chromosome is marked so it is always inactivated in the somatic cells of females. In marsupials, X inactivation is not random; the maternal X chromosome is always active.

The Imprinting of Genes and Chromosomes Is a Molecular Marking Process That Involves DNA Methylation

As we have seen, genomic imprinting must involve a marking process. A particular gene or chromosome must be marked differently



Establishment of the imprint
In this example, imprinting occurs during gametogenesis in the *Igf2* gene, which exists in the *Igf2* allele from the male and the *Igf2*$^-$ allele from the female. This imprinting occurs so that only the paternal allele is expressed.

Maintenance of the imprint
After fertilization, the imprint pattern is maintained throughout development. In this example, the maternal *Igf2*$^-$ allele will not be expressed in the somatic cells. Note that the offspring on the left is a female and the one on the right is a male; both are normal in size.

Erasure and reestablishment
During gametogenesis, the imprint is erased. The female mouse produces eggs in which the gene is silenced. The male produces sperm in which the gene can be transcribed into mRNA.

FIGURE 6.10 Genomic imprinting during gametogenesis. This example involves a mouse gene *Igf2*, which is found in two alleles designated *Igf2* and *Igf2*$^-$. The left side shows a female mouse that was produced from a sperm carrying the *Igf2* allele and an egg carrying the *Igf2*$^-$ allele. In the somatic cells of this female animal, the *Igf2* allele is active. However, when this female produces eggs, both alleles are transcriptionally inactive when they are transmitted to offspring. The right side of this figure shows a male mouse that was also produced from a sperm carrying the *Igf2* allele and an egg carrying the *Igf2*$^-$ allele. In the somatic cells of this male animal, the *Igf2* allele is active. However, the sperm from this male contains either a functionally active *Igf2* allele or a functionally defective *Igf2*$^-$ allele.

Concept check: Explain why the erasure phase of imprinting is necessary in eggs.

during spermatogenesis versus oogenesis. After fertilization takes place, this differential marking affects the expression of particular genes. What is the molecular explanation for genomic imprinting? As discussed in Chapter 17 (see Figure 17.12), **DNA methylation**—the attachment of a methyl group onto a cytosine base in DNA—is a common way that eukaryotic genes may be regulated. Research indicates that genomic imprinting involves an **imprinting control region (ICR)** that is located near the imprinted gene. The ICR contains binding sites for one or more proteins that regulate the transcription of the imprinted gene.

For most imprinted genes, methylation causes an inhibition of gene expression. Methylation could enhance the binding of proteins that inhibit transcription or inhibit the binding of proteins that enhance transcription (or both). (The relationship between methylation and gene expression is described in detail in Chapter 17.) For this reason, imprinting is usually described as a marking process that silences gene expression by preventing transcription. However, this is not always the case. Two imprinted human genes—*H19* and *Igf2*—provide an interesting example. These two genes lie close to each other on human chromosome 11 and appear to be controlled by the same ICR, which is a 52,000-bp region that lies between the *Igf2* and *H19* genes (**Figure 6.11**). It contains binding sites for proteins that regulate the transcription of the *H19* or *Igf2* genes. This ICR is highly methylated on the paternally inherited chromosome but not on the maternally inherited one.

When the ICR is not methylated, a protein called CTC-binding factor is able to bind to the ICR (Figure 6.11a). The CTC-binding factor gets its name from the observation that it binds to a region of DNA that is rich in CTC (cytosine-thymine-cytosine) sequences. The ICR contains several such CTC sequences. As described in Figure 6.11a, the binding of CTC-binding factor to the ICR has two effects. First, it prevents activator proteins from activating the *Igf2* gene, thereby shutting off this gene. In contrast, it permits activator proteins to turn on the *H19* gene.

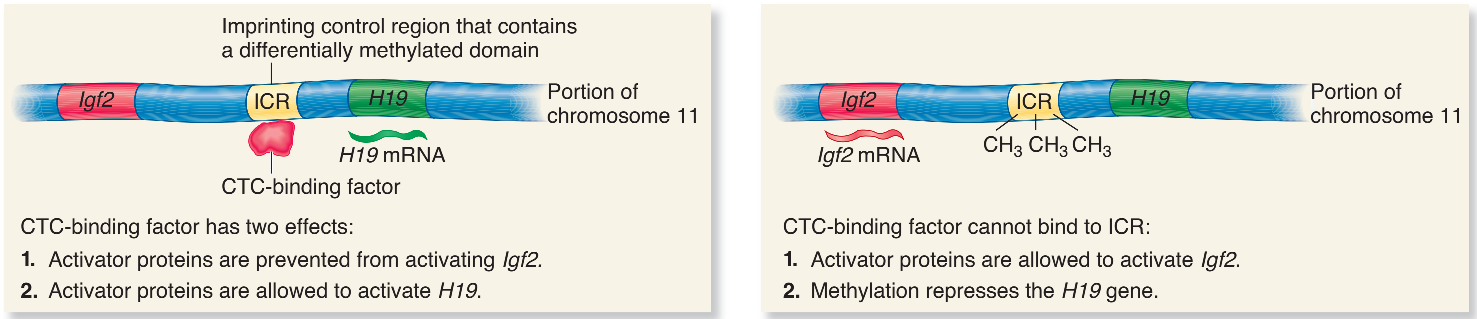
How does methylation affect the transcription of the *Igf2* and *H19* genes? When the cytosines within the ICR become methylated, the CTC-binding factor is unable to bind to the ICR (Figure 6.11b). This permits activator proteins to turn on the *Igf2* gene. The DNA methylation also causes the repression of the *H19* gene so it is not transcribed. This topic is described in greater detail in Chapter 17 (see Figure 17.16).

Imprinting from Generation to Generation Involves Maintenance, Erasure, and *de Novo* Methylation Steps

Now that we understand how methylation may affect gene transcription, let's consider the methylation process from one generation to the next. In the example shown in **Figure 6.12**, the paternally inherited allele for a particular gene is methylated, but the maternally inherited allele is not. A female (left side) and male (right side) have inherited a methylated ICR from their father and an unmethylated ICR from their mother. This pattern of imprinting is maintained in the somatic cells of both individuals. However, when the female makes gametes, the imprinting is erased (demethylated) during early oogenesis, so the female passes an unmethylated ICR to its offspring. In the male, the imprinting is also erased during early spermatogenesis, but then *de novo* (new) methylation occurs in both ICRs. Therefore, the male transmits a methylated gene to its offspring.

Imprinting Plays a Role in the Inheritance of Certain Human Genetic Diseases

Human diseases, such as Prader-Willi syndrome (PWS) and Angelman syndrome (AS), are influenced by imprinting. PWS is characterized by reduced motor function, obesity, and small hands and feet. Individuals with AS are thin and hyperactive, have unusual seizures and repetitive symmetrical muscle movements, and exhibit mental deficiencies. Most commonly, both PWS and AS involve a small deletion in human chromosome 15. If this deletion is inherited from the mother, it leads to Angelman


Imprinting control region that contains a differentially methylated domain
Igf2
ICR
H19
Portion of chromosome 11
H19 mRNA
CTC-binding factor
CTC-binding factor has two effects:
1. Activator proteins are prevented from activating Igf2.
2. Activator proteins are allowed to activate H19.
Igf2
ICR
H19
Portion of chromosome 11
Igf2 mRNA
CH_3 CH_3 CH_3
CTC-binding factor cannot bind to ICR:
1. Activator proteins are allowed to activate Igf2.
2. Methylation represses the H19 gene.

(a) Maternal chromosome **(b) Paternal chromosome**

FIGURE 6.11 **A simplified scheme for how DNA methylation at the ICR affects the expression of the *Igf2* and *H19* genes.** **(a)** The lack of methylation of the maternal chromosome causes the *Igf2* gene to be turned off and the *H19* gene to be turned on. **(b)** The methylation of the paternal chromosome has the opposite effect. The *Igf2* gene is turned on and the *H19* gene is turned off.

Concept check: *What process prevents the binding of CTC-binding factor to the ICR?*

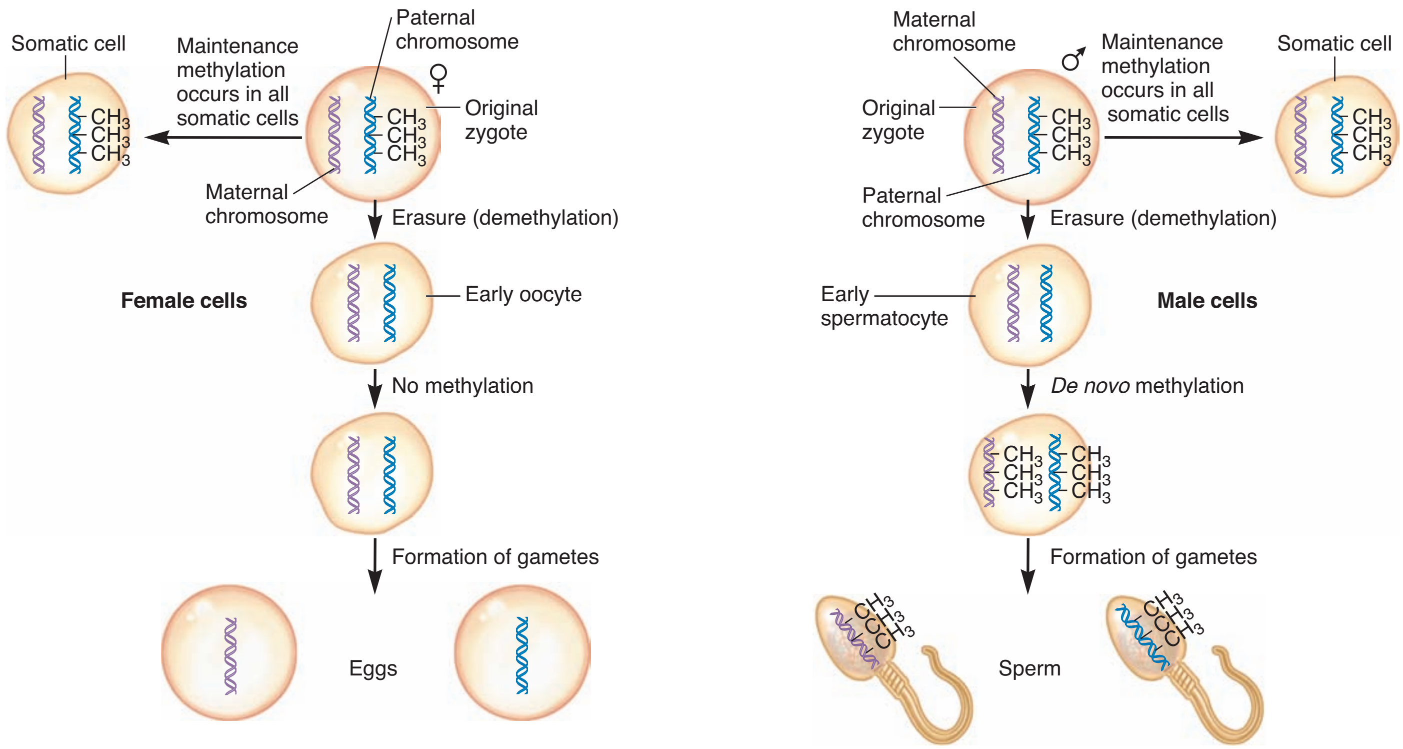

Somatic cell
Maintenance methylation occurs in all somatic cells
Paternal chromosome
♀
Original zygote
Maternal chromosome
CH3
Erasure (demethylation)
Female cells
Early oocyte
No methylation
Formation of gametes
Eggs
Maternal chromosome
♂
Maintenance methylation occurs in all somatic cells
Somatic cell
Original zygote
Paternal chromosome
Erasure (demethylation)
Early spermatocyte
Male cells
De novo methylation
Formation of gametes
Sperm

FIGURE 6.12. The pattern of ICR methylation from one generation to the next. In this example, a male and a female offspring have inherited a methylated imprinting control region (ICR) and unmethylated ICR from their father and mother, respectively. Maintenance methylation retains the imprinting in somatic cells during embryogenesis and in adulthood. Erasure (demethylation) occurs in cells that are destined to become gametes. In this example, *de novo* methylation occurs only in cells that are destined to become sperm. Haploid male gametes transmit a methylated ICR, whereas haploid female gametes transmit an unmethylated ICR.

Concept check: What is the difference between maintenance methylation and de novo methylation? In what cell types (somatic cells or germline cells), do they occur?

syndrome; if inherited from the father, it leads to Prader-Willi syndrome (**Figure 6.13**).

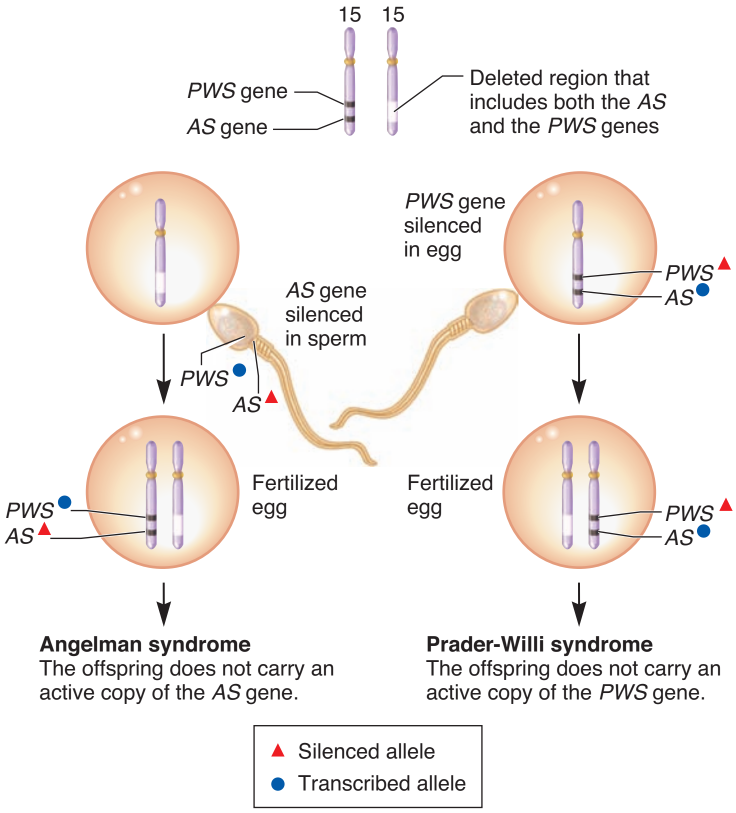
Researchers have discovered that this region of chromosome 15 contains closely linked but distinct genes that are maternally or paternally imprinted. AS results from the lack of expression of a single gene (*UBE3A*) that encodes a protein called E6-AP, which functions to transfer small ubiquitin molecules to certain proteins to target their degradation. Both copies of this gene are active in many of the body's tissues. In the brain, however, only the copy inherited from a person's mother (the maternal copy) is active. The paternal allele of *UBE3A* is silenced. Therefore, if the maternal allele is deleted, as in the left side of Figure 6.13, the individual develops AS because she or he does not have an active copy of the *UBE3A* gene.

The gene(s) responsible for PWS has not been definitively determined, although five imprinted genes in this region of chromosome 15 are known. One possible candidate involved in PWS is a gene designated *SNRPN* (for small nuclear ribonucleoprotein polypeptide N), which is part of a complex that controls RNA splicing and is necessary for the synthesis of critical proteins in the brain. The maternal allele of *SNRPN* is silenced, and only the paternal copy is active. PWS occurs when the paternal allele is deleted.

The Biological Significance of Imprinting Is Not Well Understood

Genomic imprinting is a fairly new and exciting area of research. Imprinting has been identified in many mammalian genes (**Table 6.6**). In some cases, the female alleles are transcriptionally active in the somatic cells of offspring, whereas in other cases, the male alleles are active.

The biological significance of imprinting is still a matter of much speculation. Several hypotheses have been proposed to explain the potential benefits of genomic imprinting. One example, described by David Haig, involves differences in female versus male reproductive patterns in mammals. As discussed in Chapter 25, natural selection favors types of genetic variation that confer a reproductive advantage. The likelihood that favorable variation will be passed to offspring may differ between the sexes. In many mammalian species, females may mate with multiple males, perhaps generating embryos in the same uterus fathered by different males. For males, silencing genes that inhibit embryonic growth is an advantage. The embryos of males that silence such genes grow faster than other embryos in the same uterus, making it more likely for the males to pass their genes to future generations. For females, however, rapid growth of embryos might be


15
15
PWS gene
AS gene
Deleted region that includes both the AS and the PWS genes
PWS gene silenced in egg
AS gene silenced in sperm
PWS
AS
PWS
AS
Fertilized egg
Fertilized egg
PWS
AS
PWS
AS
Angelman syndrome
The offspring does not carry an active copy of the AS gene.
Prader-Willi syndrome
The offspring does not carry an active copy of the PWS gene.
▲ Silenced allele
● Transcribed allele

ONLINE ANIMATION **FIGURE 6.13** **The role of imprinting in the development of Angelman and Prader-Willi syndromes.**

Genes → Traits A small region on chromosome 15 contains two different genes designated the *AS* gene and *PWS* gene in this figure. If a chromosome 15 deletion is inherited from the mother, Angelman syndrome occurs because the offspring does not inherit an active copy of the *AS* gene (left). Alternatively, the chromosome 15 deletion may be inherited from the father, leading to Prader-Willi syndrome. The phenotype of this syndrome occurs because the offspring does not inherit an active copy of the *PWS* gene (right).

Concept check: On the left side of this figure, explain why the offspring does not have Prader-Willi syndrome but does have Angelman syndrome.

TABLE **6.6**

Examples of Mammalian Genes and Inherited Human Diseases That Involve Imprinted Genes*

Gene	Allele Expressed	Function
WT1	Maternal	Wilms tumor-suppressor gene; suppresses cell growth
INS	Paternal	Insulin; hormone involved in cell growth and metabolism
Igf2	Paternal	Insulin-like growth factor 2; similar to insulin
Igf2R	Maternal	Receptor for insulin-like growth factor 2
H19	Maternal	Unknown
SNRPN	Paternal	Splicing factor
Gabrb	Maternal	Neurotransmitter receptor

*Researchers estimate that approximately 1–2% of human genes are subjected to genomic imprinting, but fewer than 100 have actually been demonstrated to be imprinted.

a disadvantage because it could drain too many resources from the mother. According to this scenario, the mother silences genes that cause rapid embryonic growth. From the mother's perspective, she gives all of her offspring an equal chance of survival without sapping her own strength. This makes it more likely for the female to pass her genes to future generations.

The Haig hypothesis seems to be consistent with the imprinting of several mammalian genes that are involved with growth, such as *Igf2*. Females silence this growth-enhancing gene, whereas males do not. However, several imprinted genes do not appear to play a role in embryonic development. Therefore, an understanding of the biological role(s) of genomic imprinting requires further investigation.

6.4 REVIEWING THE KEY CONCEPTS

- Epigenetic inheritance is an inheritance pattern in which a gene or chromosome is modified so as to alter gene expression, but the modification is not permanent over the course of many generations. An example is imprinting in which an offspring expresses a gene that is inherited from one parent but not both (see Figures 6.9, 6.10).
- DNA methylation at imprinting control regions is the marking process that causes imprinting (see Figures 6.11, 6.12).
- Human diseases such as Prader-Willi syndrome and Angelman syndrome are associated with genomic imprinting (see Figure 6.13, Table 6.6).

6.4 COMPREHENSION QUESTIONS

1. In mice, the copy of the *Igf2* gene that is inherited from the mother is never expressed in her offspring. This happens because the *Igf2* gene from the mother
    a. always undergoes a mutation that inactivates its function.
    b. is deleted during oogenesis.
    c. is deleted during embryonic development.
    d. is not transcribed in the somatic cells of the offspring.
2. A female mouse, *Igf2* $Igf2^-$, is crossed to a male that is also *Igf2* $Igf2^-$. The expected outcome of the offspring for this cross is
    a. all normal.
    b. all dwarf.
    c. 1 normal : 1 dwarf.
    d. 3 normal : 1 dwarf.
3. The marking process for genomic imprinting initially occurs during
    a. gametogenesis.
    b. fertilization.
    c. embryonic development.
    d. adulthood.
4. A female born with Angelman syndrome carries a deletion in the AS gene (i.e., the *UBE3A gene*). Which parent transmitted the deletion to her?
    a. Her father
    b. Her mother
    c. Either her mother or father

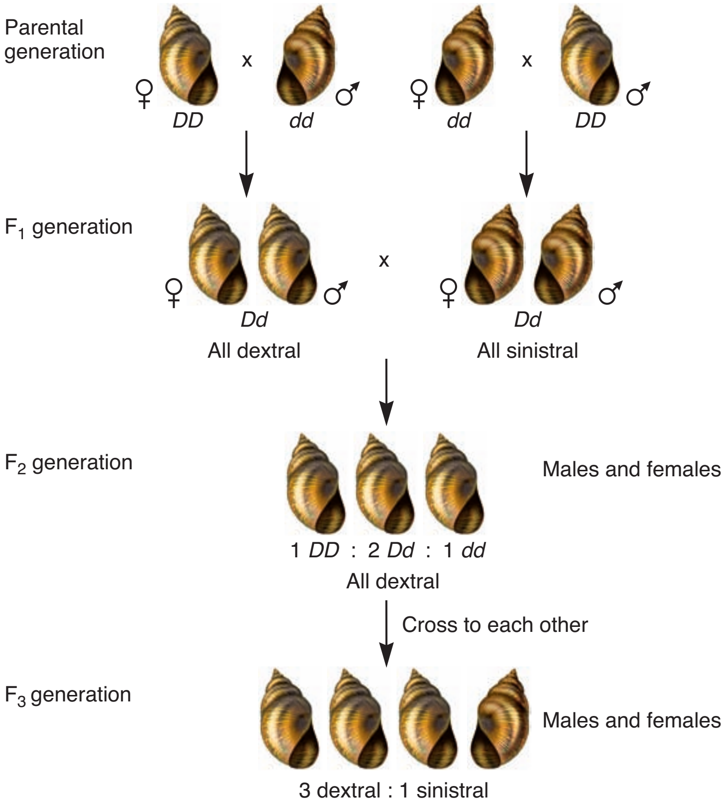
6.5 MATERNAL EFFECT

Maternal effect refers to an inheritance pattern for certain nuclear genes in which the genotype of the mother directly determines the phenotype of her offspring. (Note that maternal effect should not be confused with maternal inheritance.) Surprisingly, for maternal effect genes, the genotypes of the father and offspring themselves do not affect the phenotype of the offspring. We will see that this phenomenon is explained by the accumulation of gene products that the mother provides to her developing eggs.

The Genotype of the Mother Determines the Phenotype of the Offspring for Maternal Effect Genes

The first example of a maternal effect gene was studied in the 1920s by Arthur Boycott and involved morphological features of the water snail, *Lymnaea peregra*. In this species, the shell and internal organs can be arranged in either a right-handed (dextral) or left-handed (sinistral) direction. The dextral orientation is more common and is dominant to the sinistral orientation. **Figure 6.14** describes the results of a genetic analysis carried out by Boycott. In this experiment, he began with two different true-breeding strains of snails with either a dextral or sinistral morphology. Many combinations of crosses produced results that could not be explained by a Mendelian pattern of inheritance. When a dextral female (*DD*) was crossed to a sinistral male (*dd*), all F_1 offspring were dextral. However, in the **reciprocal cross,** where a sinistral female (*dd)* was crossed to a dextral male (*DD*), all F_1 offspring were sinistral. Taken together, these results contradict a Mendelian pattern of inheritance.

How can we explain the unusual results obtained in Figure 6.14? Alfred Sturtevant proposed the idea that snail coiling is due to a maternal effect gene that exists as a dextral (*D*) or sinistral (*d*) allele. His conclusions were drawn from the inheritance patterns of the F_2 and F_3 generations. In this experiment, the genotype of the F_1 generation is expected to be heterozygous (*Dd*). When these F_1 individuals were crossed to each other, a genotypic ratio of 1 *DD* : 2 *Dd* : 1 *dd* is predicted for the F_2 generation. Because the *D* allele is dominant to the *d* allele, a 3:1 phenotypic ratio of dextral to sinistral snails should be produced according to a Mendelian pattern of inheritance. Instead of this predicted phenotypic ratio, however, the F_2 generation was composed of all dextral snails. This incongruity with Mendelian inheritance is due to the maternal effect. The phenotype of the offspring depended solely on the genotype of the mother. The F_1 mothers were *Dd*. The *D* allele in the mothers is dominant to the *d* allele and caused the offspring to be dextral, even if the offspring's genotype was *dd*. When the members of the F_2 generation were crossed, the F_3 generation exhibited a 3:1 ratio of dextral to sinistral snails. This ratio corresponds to the genotypes of the F_2 females, which were the mothers of the F_3 generation. The ratio of F_2 females was 1 *DD* : 2 *Dd* : 1 *dd*. The *DD* and *Dd* females produced dextral offspring, whereas the *dd* females produced sinistral offspring. This explains the 3:1 ratio of dextral and sinistral offspring in the F_3 generation.



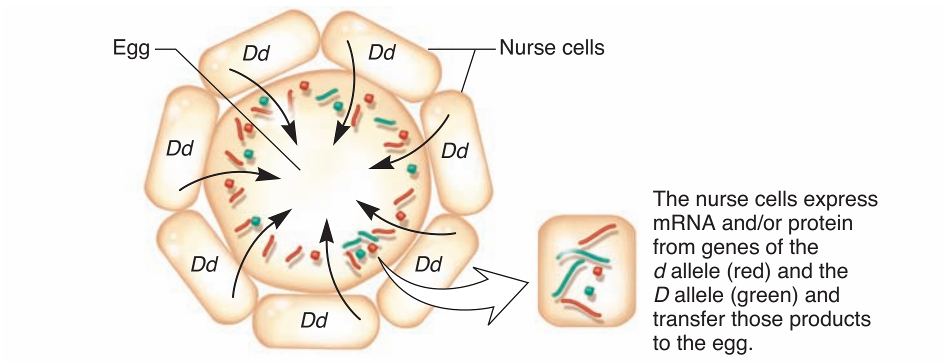
INTERACTIVE EXERCISE

FIGURE 6.14 Experiment showing the inheritance pattern of snail coiling. In this experiment, *D* (dextral) is dominant to *d* (sinistral). The genotype of the mother determines the phenotype of the offspring. This phenomenon is known as the maternal effect. In this case, a *DD* or *Dd* mother produces dextral offspring, and a *dd* mother produces sinistral offspring. The genotypes of the father and offspring do not affect the offspring's phenotype.

Concept check: Explain why all of the offspring in the F_2 generation are dextral even though some of them are dd.

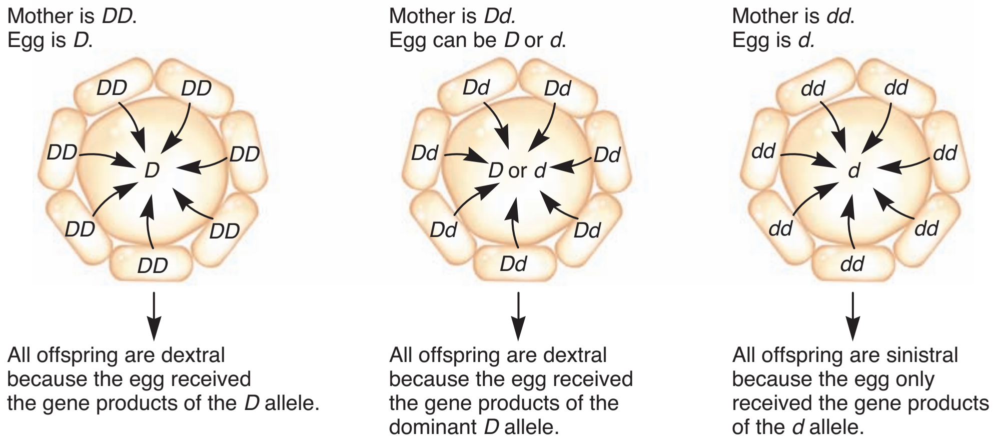
Female Gametes Receive Gene Products from the Mother That Affect Early Developmental Stages of the Embryo

At the molecular and cellular level, the non-Mendelian inheritance pattern of maternal effect genes can be explained by the process of oogenesis in female animals (**Figure 6.15a**). As an animal oocyte (egg) matures, many surrounding maternal cells called nurse cells provide the egg with nutrients and other materials. In Figure 6.15a, a female is heterozygous for the snail-coiling maternal effect gene, with the alleles designated *D* and *d*. Depending on the outcome of meiosis, the haploid egg may receive the *D* allele or the *d* allele, but not both. The surrounding nurse cells, however, produce both *D* and *d* gene products (mRNA and/or proteins). These gene products are then transported into the egg. As shown here, the egg has received both the *D* allele gene products and the *d* allele gene products. These gene products persist for a significant time after the egg has been fertilized and begins its embryonic development. In this way, the



The nurse cells express mRNA and/or protein from genes of the *d* allele (red) and the *D* allele (green) and transfer those products to the egg.

(a) Transfer of gene products from nurse cells to egg



Mother is *DD*.
Egg is *D*.

All offspring are dextral because the egg received the gene products of the *D* allele.

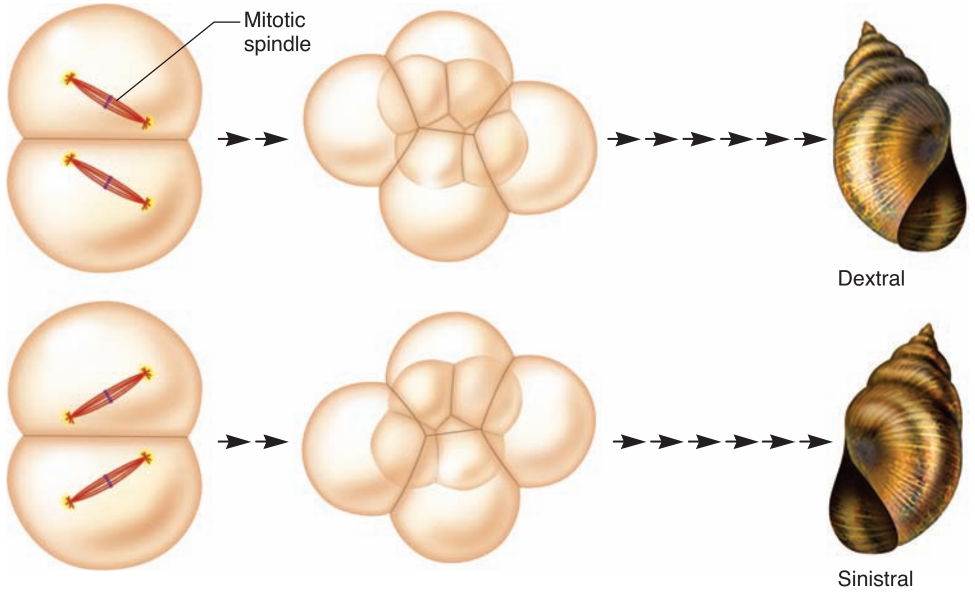
Mother is *Dd*.
Egg can be *D* or *d*.

All offspring are dextral because the egg received the gene products of the dominant *D* allele.

Mother is *dd*.
Egg is *d*.

All offspring are sinistral because the egg only received the gene products of the *d* allele.

(b) Maternal effect in snail coiling



(c) An explanation of coiling direction at the cellular level

FIGURE 6.15 **The mechanism of maternal effect in snail coiling.** **(a)** Transfer of gene products from nurse cells to an egg. The nurse cells are heterozygous (*Dd*). Both the *D* and *d* alleles are activated in the nurse cells to produce *D* and *d* gene products (mRNA or proteins, or both). These products are transported into the cytoplasm of the egg, where they accumulate to significant amounts. **(b)** Explanation of the maternal effect in snail coiling. **(c)** The direction of snail coiling is determined by differences in the cleavage planes during early embryonic development.

Genes → Traits If the nurse cells are *DD* or *Dd*, they will transfer the *D* gene product to the egg, thereby causing the offspring to be dextral. If the nurse cells are *dd*, only the *d* gene product is transferred to the egg, so the resulting offspring will be sinistral.

Concept check: *If a mother is heterozygous, Dd, which gene products will the egg cell receive?*

gene products of the nurse cells, which reflect the genotype of the mother, influence the early developmental stages of the embryo.

Now that we understand the relationship between oogenesis and maternal effect genes, let's reconsider the topic of snail coiling. As shown in **Figure 6.15b**, a female snail that is *DD* transmits only the *D* gene products to the egg. During the early stages of embryonic development, these gene products cause the egg cleavage to occur in a way that promotes a right-handed body plan. A heterozygous female transmits both *D* and *d* gene products. Because the *D* allele is dominant, the maternal effect also causes a right-handed body plan. Finally, a *dd* mother contributes only *d* gene products that promote a left-handed body plan, even if the egg is fertilized by a sperm carrying a *D* allele. The sperm's genotype is irrelevant, because the expression of the sperm's gene would occur too late. The origin of dextral and sinistral coiling can be traced to the orientation of the mitotic spindle at the two- to four-cell stage of embryonic development. The dextral and sinistral snails develop as mirror images of each other (**Figure 6.15c**).

Since these initial studies, researchers have found that maternal effect genes encode proteins that are important in the early steps of embryogenesis. The accumulation of maternal gene products in the egg allows embryogenesis to proceed quickly after fertilization. Maternal effect genes often play a role in cell division, cleavage pattern, and body axis orientation. Therefore, defective alleles in maternal effect genes tend to have a dramatic effect on the phenotype of the individual, altering major features of morphology, often with dire consequences.

Our understanding of maternal effect genes has been greatly aided by their identification in experimental organisms such as *Drosophila melanogaster*. In such organisms with a short generation time, geneticists have successfully searched for mutant alleles that prevent the normal process of embryonic development. In *Drosophila,* geneticists have identified several maternal effect genes that have profound effects on the early stages of development. The pattern of development of a *Drosophila* embryo occurs along axes, such as the anteroposterior axis and the dorsoventral axis. The proper development of each axis requires a distinct set of maternal gene products. For example, the maternal effect gene called *bicoid* produces a gene product that accumulates in a region of the egg that eventually becomes anterior structures in the developing embryo. Mutant alleles of maternal effect genes often lead to abnormalities in the anteroposterior or the dorsoventral pattern of development. More recently, several maternal effect genes have been identified in mice and humans that are required for proper embryonic development. Chapter 24 examines the relationships among the actions of several maternal effect genes during embryonic development.

6.5 REVIEWING THE KEY CONCEPTS

- Maternal effect is an inheritance pattern in which the genotype of the mother determines the phenotype of the offspring (see Figure 6.14).
- Maternal effect occurs because the gene products of maternal effect genes are transferred from nurse cells to the oocyte. These gene products affect early stages of development (see Figure 6.15).

6.5 COMPREHENSION QUESTIONS

1. A female that coils to the left has offspring that coil to the right. What are the genotypes of the mother of these offspring and the maternal grandmother, respectively?
   a. *dd*, *DD*
   b. *Dd*, *Dd*
   c. *dd*, *Dd*
   d. *Dd*, *dd*
2. What is the molecular–cellular explanation for maternal effect?
   a. The father's gene is silenced at fertilization.
   b. During oogenesis, nurse cells transfer gene products to the oocyte.
   c. The gene products from nurse cells are needed during the very early stages of development.
   d. Both b and c are correct.

KEY TERMS

Page 110. extranuclear inheritance, cytoplasmic inheritance
Page 111. nucleoid, chloroplast DNA (cpDNA), nuclear genes, maternal inheritance
Page 112. heteroplasmy
Page 114. heterogamous
Page 115. mitochondrial DNA (mtDNA), paternal leakage
Page 116. endosymbiosis, endosymbiosis theory
Page 118. epigenetic inheritance, genomic imprinting, imprinting, monoallelic expression
Page 120. DNA methylation, imprinting control region (ICR)
Page 123. maternal effect, reciprocal cross

CHAPTER SUMMARY

- Non-Mendelian inheritance refers to inheritance patterns that do not obey Mendel's laws of inheritance.

6.1 Extranuclear Inheritance: Chloroplasts

- Extranuclear inheritance involves the inheritance of genes that are found in chloroplasts or mitochondria.
- Chloroplasts carry circular chromosomes in a nucleoid region. These circular chromosomes contain many genes but far fewer compared with the number in the cell nucleus (see Figures 6.1, 6.2, Table 6.1).
- Maternal inheritance occurs when organelles, such as chloroplasts, are transmitted via the egg (see Figure 6.3).

- Heteroplasmy for chloroplasts can result in a variegated phenotype (see Figure 6.4).
- In the alga *Chlamydomonas,* chloroplasts are transmitted from the *mt*$^+$ parent (see Figure 6.5).
- The transmission patterns of chloroplasts vary among different species (see Table 6.2).

6.2 Extranuclear Inheritance: Mitochondria

- Mitochondria carry circular chromosomes in a nucleoid region (see Figures 6.6, 6.7, Table 6.3).
- Mitochondria are usually transmitted by a maternal inheritance pattern, but transmission patterns vary among different species (see Table 6.4).
- Many human diseases are caused by mutations in mitochondrial DNA (see Table 6.5).

6.3 Theory of Endosymbiosis

- Chloroplasts and mitochondria were derived from an ancient endosymbiotic relationship (see Figure 6.8).

6.4 Epigenetics: Imprinting

- Epigenetic inheritance is an inheritance pattern in which a gene or chromosome is modified so as to alter gene expression, but the modification is not permanent over the course of many generations. An example is imprinting in which an offspring expresses a gene that is inherited from one parent but not both (see Figures 6.9, 6.10).
- DNA methylation at imprinting control regions is the marking process that causes imprinting (see Figures 6.11, 6.12).
- Human diseases such as Prader-Willi syndrome and Angelman syndrome are associated with genomic imprinting (see Figure 6.13, Table 6.6).

6.5 Maternal Effect

- Maternal effect is an inheritance pattern in which the genotype of the mother determines the phenotype of the offspring (see Figure 6.14).
- Maternal effect occurs because the gene products of maternal effect genes are transferred from nurse cells to the oocyte. These gene products affect early stages of development (see Figure 6.15).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Suppose that you are a horticulturist who has recently identified an interesting plant with variegated leaves. How would you determine if this trait is the result of nuclear or cytoplasmic inheritance?

Answer: Make crosses and reciprocal crosses involving normal and variegated strains. In many species, chloroplast genomes are inherited maternally, although this is not always the case. In addition, a significant percentage of paternal leakage may occur. Nevertheless, when reciprocal crosses yield different outcomes, an organellar mode of inheritance is possibly at work.

S2. Discuss the two general ways that mitochondrial diseases may occur.

Answer: In some cases, mitochondrial mutations that cause human disease are transmitted from mother to offspring. This follows a maternal inheritance pattern. A second way that mitochondrial diseases occur is via somatic mutations. Mitochondrial mutations may occur in somatic cells and accumulate as a person ages. An important factor in mitochondrial disease is heteroplasmy. To cause disease that affects a particular cell or tissue, the ratio of mutant to normal mitochondria must exceed a certain threshold value before disease symptoms are observed.

S3. An individual with Angelman syndrome produced an offspring with Prader-Willi syndrome. Why does this occur? What are the sexes of the parent with Angelman syndrome and the offspring with Prader-Willi syndrome?

Answer: These two different syndromes are most commonly caused by a small deletion in chromosome 15. In addition, genomic imprinting plays a role because genes in this deleted region are differentially imprinted, depending on sex. If this deletion is inherited from the paternal parent, the offspring develops Prader-Willi syndrome. Therefore, in this problem, the person with Angelman syndrome must have been a male because he produced a child with Prader-Willi syndrome. The child could be either a male or female.

S4. Our understanding of maternal effect genes has been greatly aided by their identification in experimental organisms such as *Drosophila melanogaster*. In experimental organisms with a short generation time, geneticists have successfully searched for mutant alleles that prevent the normal process of embryonic development. In many cases, the offspring die at early embryonic or larval stages. These are called maternal effect lethal alleles. How would a researcher identify a mutation that produced a recessive maternal effect lethal allele?

Answer: A maternal effect lethal allele can be identified when a phenotypically normal mother produces only offspring with gross developmental abnormalities. For example, let's call the normal allele *N* and the maternal effect lethal allele *n*. A cross between two flies that are heterozygous for a maternal effect lethal allele produces 1/4 of the offspring with a homozygous genotype, *nn*. These flies are viable because of the maternal effect. Their mother is *Nn* and provides the *n* egg with a sufficient amount of *N* gene product so the *nn* flies develop properly. However, homozygous *nn* females cannot provide their eggs with any normal gene product. Therefore, all of their offspring would be abnormal and die during early stages.

Conceptual Questions

C1. What is extranuclear inheritance? Describe three examples.

C2. Among different species, does extranuclear inheritance always follow a maternal inheritance pattern? Why or why not?

C3. Extranuclear inheritance often correlates with maternal inheritance. Even so, paternal leakage can occur. What is paternal leakage? If a cross produced 200 offspring and the level of mitochondrial paternal leakage was 3%, how many offspring would be expected to contain paternal mitochondria?

C4. Discuss the structure and organization of the mitochondrial and chloroplast genomes. How large are they, how many genes do they contain, and how many copies of the genome are found in each organelle?

C5. Explain the likely evolutionary origin of chloroplast and mitochondrial genomes. How have the sizes of the chloroplast and mitochondrial genomes changed since their origin? How did this occur?

C6. Which of the following traits or diseases are determined by nuclear genes?

A. Snail coiling pattern

B. Prader-Willi syndrome

C. Streptomycin resistance in *Chlamydomonas*

D. Leber hereditary optic neuropathy

C7. Acute murine leukemia virus (AMLV) causes leukemia in mice. This virus is easily passed from mother to offspring through the mother's milk. (*Note:* Even though newborn offspring acquire the virus, they may not develop leukemia until much later in life. Testing can determine if an animal carries the virus.) Describe how the formation of leukemia via AMLV resembles a maternal inheritance pattern. How could you determine that this form of leukemia is not caused by extranuclear inheritance?

C8. Describe how a biparental pattern of extranuclear inheritance would resemble a Mendelian pattern of inheritance for a particular gene. How would they differ?

C9. According to the endosymbiosis theory, chloroplasts and mitochondria are derived from bacteria that took up residence within eukaryotic cells. At one time, prior to being taken up by eukaryotic cells, these bacteria were free-living organisms. However, we cannot take a chloroplast or mitochondrion out of a living eukaryotic cell and get it to survive and replicate on its own. Why not?

C10. Define the term *genomic imprinting* and describe two examples.

C11. When does the erasure and reestablishment phase of genomic imprinting occur? Explain why it is necessary to erase an imprint and then reestablish it according to the sex of the parent.

C12. In what types of cells would you expect *de novo* methylation to occur? In what cell types would it not occur?

C13. Genes that cause Prader-Willi syndrome and Angelman syndrome are closely linked along chromosome 15. Although people with these syndromes do not usually reproduce, let's suppose that a couple produces two children with Angelman syndrome. The oldest child (named Pat) grows up and has two children with Prader-Willi syndrome. The second child (named Robin) grows up and has one child with Angelman syndrome.

A. Are Pat and Robin's parents both normal or does one of them have Angelman syndrome or Prader-Willi syndrome? If one of them has a disorder, explain why it is the mother or the father.

B. What are the sexes of Pat and Robin? Explain.

C14. How is the process of random X inactivation in mammals, which is described in Chapter 4, similar to genomic imprinting? How is it different?

C15. Describe the inheritance pattern of maternal effect genes. Explain how the maternal effect occurs at the cellular level. What are the expected functional roles of the proteins that are encoded by maternal effect genes?

C16. A maternal effect gene exists in a dominant *N* (normal) allele and a recessive *n* (abnormal) allele. What would be the ratios of genotypes and phenotypes for the offspring of the following crosses?

A. *nn* female × *NN* male

B. *NN* female × *nn* male

C. *Nn* female × *Nn* male

C17. A *Drosophila* embryo dies during early embryogenesis due to a recessive maternal effect allele called *bicoid*. The wild-type allele is designated *bicoid*$^+$. What are the genotypes and phenotypes of the embryo's mother and maternal grandparents?

C18. For Mendelian traits, the nuclear genotype (i.e., the alleles found on chromosomes in the cell nucleus) directly influences an offspring's traits. In contrast, for non-Mendelian inheritance patterns, the offspring's phenotype cannot be reliably predicted solely from its genotype. For the following traits, what do you need to know to predict the phenotypic outcome?

A. Dwarfism due to a mutant *Igf2* allele

B. Snail coiling direction

C. Leber hereditary optic neuropathy

C19. Suppose a maternal effect gene exists as a normal dominant allele and an abnormal recessive allele. A mother who is phenotypically abnormal produces all normal offspring. Explain the genotype of the mother.

C20. Suppose that a gene affects the anterior morphology in house flies and is inherited as a maternal effect gene. The gene exists in a normal allele, *H*, and a recessive allele, *h*, which causes a small head. A female fly with a normal head is mated to a true-breeding male with a small head. All of the offspring have small heads. What are the genotypes of the mother and offspring? Explain your answer.

C21. Explain why maternal effect genes exert their effects during the early stages of development.

Application and Experimental Questions

E1. A variegated trait in plants is analyzed using reciprocal crosses. The following results are obtained:

Variegated female × Normal male	Normal female × Variegated male
↓	↓
1024 variegated + 52 normal	1113 normal + 61 variegated

Explain this pattern of inheritance.

E2. Ruth Sager and her colleagues discovered that the mode of inheritance of streptomycin resistance in *Chlamydomonas* could be altered if the mt^+ cells were exposed to UV irradiation prior to mating. This exposure dramatically increased the frequency of biparental inheritance. What would be the expected outcome of a cross between an $mt^+\ sm^r$ and an $mt^-\ sm^s$ strain in the absence of UV irradiation? How would the results differ if the mt^+ strain was exposed to UV light?

E3. Two male mice, which we will call male A and male B, are both phenotypically normal. Male A was from a litter that contained half phenotypically normal mice and half dwarf mice. The mother of male A was known to be homozygous for the normal *Igf2* allele. Male B was from a litter of eight mice that were all phenotypically normal. The parents of male B were a phenotypically normal male and a dwarf female. Male A and male B were put into a cage with two female mice that we will call female A and female B. Female A is dwarf, and female B is phenotypically normal. The parents of these two females were unknown, although it was known that they were from the same litter. The mice were allowed to mate with each other, and the following data were obtained:

Female A gave birth to three dwarf babies and four normal babies.

Female B gave birth to four normal babies and two dwarf babies.

Which male(s) mated with female A and female B? Explain.

E4. Figure 6.14 describes an example of a maternal effect gene. Explain how Sturtevant deduced a maternal effect gene based on the F_2 and F_3 generations.

E5. Chapter 19 describes three blotting methods (i.e., Southern blotting, Northern blotting, and Western blotting) that are used to detect specific genes and gene products. Southern blotting detects DNA, Northern blotting detects RNA, and Western blotting detects proteins. Suppose that a female fruit fly is heterozygous for a maternal effect gene, which we will call gene *B*. The female is *Bb*. The normal allele, *B*, encodes a functional mRNA that is 550 nucleotides long. A recessive allele, *b*, encodes a shorter mRNA that is 375 nucleotides long. (Allele *b* is due to a deletion within this gene.) How could you use one or more of these techniques to show that nurse cells transfer gene products from gene *B* to developing eggs? You may assume that you can dissect the ovaries of fruit flies and isolate eggs separately from nurse cells. In your answer, describe your expected results.

E6. As a hypothetical example, a trait in mice results in mice with long tails. You initially have a true-breeding strain with normal tails and a true-breeding strain with long tails. You then make the following types of crosses:

Cross 1: When true-breeding females with normal tails are crossed to true-breeding males with long tails, all F_1 offspring have long tails.

Cross 2: When true-breeding females with long tails are crossed to true-breeding males with normal tails, all F_1 offspring have normal tails.

Cross 3: When F_1 females from cross 1 are crossed to true-breeding males with normal tails, all offspring have normal tails.

Cross 4: When F_1 males from cross 1 are crossed to true-breeding females with long tails, half of the offspring have normal tails and half have long tails.

Explain the pattern of inheritance of this trait.

E7. You have a female snail that coils to the right, but you do not know its genotype. You may assume that right coiling (*D*) is dominant to left coiling (*d*). You also have male snails at your disposal of known genotype. How would you determine the genotype of this female snail? In your answer, describe your expected results depending on whether the female is *DD*, *Dd*, or *dd*.

Questions for Student Discussion/Collaboration

1. During the course of evolution, most organellar genes have been transferred from the chloroplast or mitochondrial genome to the nuclear genome. Discuss possible reasons why this may have occurred. In other words, what are possible selective advantages to having genes in the nucleus?

2. Recessive maternal effect genes are identified in flies (for example) when a phenotypically normal mother cannot produce any normal offspring. Because all of the offspring are dead, this female fly cannot be used to produce a strain of heterozygous flies that could be used in future studies. How would you identify heterozygous individuals that are carrying a recessive maternal effect allele? How would you maintain this strain of flies in a laboratory over many generations?

Answers to Comprehension Questions

6.1: b, d, a

6.2: d, a, c

6.3: b

6.4: d, c, a, b

6.5: d, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

7

Crossing over during meiosis. *This event provides a way to reassort the alleles of genes that are located on the same chromosome.*

CHAPTER OUTLINE

7.1 Overview of Linkage

7.2 Relationship Between Linkage and Crossing Over

7.3 Genetic Mapping in Plants and Animals

7.4 Mitotic Recombination

GENETIC LINKAGE AND MAPPING IN EUKARYOTES

In Chapter 3, we were introduced to Mendel's laws of inheritance. According to these principles, we expect that two different genes will segregate and independently assort themselves during the process that creates gametes. After Mendel's work was rediscovered at the turn of the twentieth century, chromosomes were identified as the cellular structures that carry genes. The chromosome theory of inheritance explained how the transmission patterns of chromosomes are responsible for the passage of genes from parents to offspring.

When geneticists first realized that chromosomes contain the genetic material, they began to suspect that a conflict might sometimes occur between the law of independent assortment of genes and the behavior of chromosomes during meiosis. In particular, geneticists assumed that each species of organism must contain thousands of different genes, yet cytological studies revealed that most species have at most a few dozen chromosomes. Therefore, it seemed likely, and turned out to be true, that each chromosome carries many hundreds or even thousands of different genes. The transmission of genes located close to each other on the same chromosome violates the law of independent assortment.

In this chapter, we will consider the pattern of inheritance that occurs when different genes are situated on the same chromosome. In addition, we will briefly explore how the data from genetic crosses are used to construct a **genetic map**—a diagram that describes the order of genes along a chromosome. Newer strategies for gene mapping are described in Chapter 21. However, an understanding of traditional mapping studies, as described in this chapter, will strengthen our appreciation for these newer molecular approaches. More importantly, traditional mapping studies further illustrate how the location of two or more genes on the same chromosome can affect the transmission patterns from parents to offspring.

7.1 OVERVIEW OF LINKAGE

In eukaryotic species, each linear chromosome contains a very long segment of DNA. A chromosome contains many individual functional units—called genes—that influence an organism's traits. A typical chromosome is expected to contain many hundreds or perhaps a few thousand different genes. The term **synteny** means that two or more genes are located on the same chromosome. Genes that are syntenic are physically linked to each other, because each eukaryotic chromosome contains a single, continuous, linear molecule of DNA. **Genetic linkage,** or simply **linkage,** is the phenomenon in which genes that are close together on the same chromosome tend to be transmitted as a unit. For this reason, linkage has an influence on inheritance patterns.

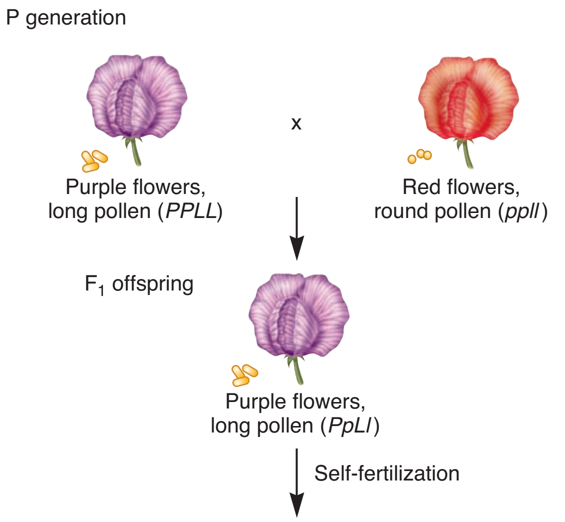
Chromosomes are sometimes called **linkage groups,** because a chromosome contains a group of genes that are physically linked together. In each species, the number of linkage groups equals the number of chromosome types. For example, human somatic cells have 46 chromosomes, which are composed of 22 types of autosomes that come in pairs plus one pair of sex chromosomes, the X and Y. Therefore, humans have 22 autosomal linkage groups, an X chromosome linkage group, and males have a Y chromosome linkage group. In addition, the human mitochondrial genome is another linkage group.

Geneticists are often interested in the transmission of two or more characters in a genetic cross. When a geneticist follows the variants of two different characters in a cross, this is called a **dihybrid cross;** when three characters are followed, it is a **trihybrid cross;** and so on. The outcome of a dihybrid or trihybrid cross depends on whether or not the genes are linked to each other along the same chromosome. Next, we will examine how linkage affects the transmission patterns of two traits. Later sections will examine crosses involving three characters.

Bateson and Punnett Discovered Two Traits That Did Not Assort Independently

An early study indicating that some traits may not assort independently was carried out by William Bateson and Reginald Punnett in 1905. According to Mendel's law of independent assortment, a dihybrid cross between two individuals, heterozygous for two genes, should yield a 9:3:3:1 phenotypic ratio among the offspring. However, a surprising result occurred when Bateson and Punnett conducted a cross in the sweet pea involving two different traits: flower color and pollen shape.

As seen in **Figure 7.1**, they began by crossing a true-breeding strain with purple flowers (*PP*) and long pollen (*LL*) to a strain with red flowers (*pp*) and round pollen (*ll*). This yielded an F_1 generation of plants that all had purple flowers and long pollen (*PpLl*). An unexpected result came from the F_2 generation. Even though the F_2 generation had four different phenotypic categories, the observed numbers of offspring did not conform to the expected 9:3:3:1 ratio. Bateson and Punnett found that the F_2 generation had a much greater proportion of the two phenotypes found in the parental generation—purple flowers with long pollen and red flowers with round pollen. Therefore, they suggested that the transmission of these two traits from the parental generation to the F_2 generation was somehow coupled and not easily assorted in an independent manner. However, Bateson and Punnett did not realize that this coupling was due to the linkage of the flower color gene and the pollen shape gene on the same chromosome.


P generation
x
Purple flowers, long pollen (*PPLL*)
Red flowers, round pollen (*ppll*)
F_1 offspring
Purple flowers, long pollen (*PpLl*)
Self-fertilization

F_2 offspring	Observed number	Ratio	Expected number	Ratio
Purple flowers, long pollen	296	15.6	240	9
Purple flowers, round pollen	19	1.0	80	3
Red flowers, long pollen	27	1.4	80	3
Red flowers, round pollen	85	4.5	27	1

FIGURE 7.1 **An experiment of Bateson and Punnett with sweet peas, showing that independent assortment does not always occur.**

Note: The expected numbers are rounded to the nearest whole number.

Genes → Traits Two genes that govern flower color and pollen shape are found on the same chromosome. Therefore, the offspring tend to inherit the parental combinations of alleles (*PL* or *pl*). Due to occasional crossing over, a lower percentage of offspring inherit recombinant combinations of alleles (*Pl* or *pL*).

Concept check: Which types of offspring are found in excess based on Mendel's law of independent assortment?

7.1 REVIEWING THE KEY CONCEPTS

- Linkage has two related meanings. First, it refers to genes that are located on the same chromosome. Second, linkage also means that the alleles of two or more genes tend to be transmitted as a unit because they are relatively close on the same chromosome.
- Bateson and Punnett discovered the first example of genetic linkage in sweet peas (see Figure 7.1).

7.1 COMPREHENSION QUESTIONS

1. Genetic linkage occurs because
   a. genes that are on the same chromosome may affect the same trait.
   b. genes that are close together on the same chromosome tend to be transmitted together to offspring.
   c. genes that are on different chromosomes are independently assorted.
   d. none of the above.

2. In the experiment by Bateson and Punnett, which of the following observations suggested linkage in the sweet pea?
   a. A 9:3:3:1 ratio was observed in the F_2 offspring.
   b. A 9:3:3:1 ratio was not observed in the F_2 offspring.
   c. An unusually high number of F_2 offspring had phenotypes of the parental generation.
   d. Both b and c suggested linkage.

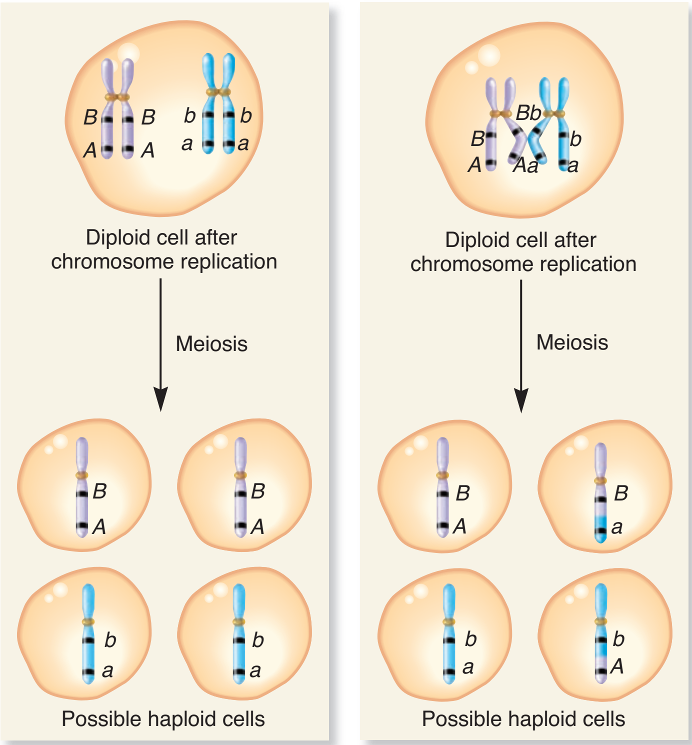
7.2 RELATIONSHIP BETWEEN LINKAGE AND CROSSING OVER

Even though the alleles for different genes may be linked along the same chromosome, the linkage can be altered during meiosis. In diploid eukaryotic species, homologous chromosomes can exchange pieces with each other, a phenomenon called **crossing over.** This event occurs during prophase of meiosis I. As discussed in Chapter 2, the replicated chromosomes, known as sister chromatids, associate with the homologous sister chromatids to form a structure known as a **bivalent.** A bivalent is composed of two pairs of sister chromatids. In prophase of meiosis I, it is common for a sister chromatid of one pair to cross over with a sister chromatid from the homologous pair (refer back to Figure 2.10). In this section, we will consider how crossing over affects the pattern of inheritance for genes linked on the same chromosome. In Chapter 13, we will consider the molecular events that cause crossing over to occur.

Crossing Over May Produce Recombinant Genotypes

Figure 7.2 considers meiosis when two genes are linked on the same chromosome. One of the parental chromosomes carries the *A* and *B* alleles, whereas the homolog carries the *a* and *b* alleles. In Figure 7.2a, no crossing over has occurred. Therefore, the resulting haploid cells contain the same combination of alleles as the original chromosomes. In this case, two haploid cells carry the dominant *A* and *B* alleles, and the other two carry the recessive *a* and *b* alleles. These are called **nonrecombinant** cells, or **parental** cells. The arrangement of linked alleles has not been altered from those found in the original cell.

In contrast, Figure 7.2b illustrates what can happen when crossing over occurs. Two of the haploid cells contain combinations of alleles, namely *A* and *b* or *a* and *B*, which differ from those in the original chromosomes. In these two cells, the grouping of linked alleles has changed. An event such as this, leading to a new combination of alleles, is known as **genetic recombination.** The haploid cells carrying the *A* and *b*, or the *a* and *B*, alleles are called **recombinant** cells, or **nonparental** cells. Likewise, if such haploid cells were gametes that participated in fertilization, the resulting offspring are called recombinant offspring, or nonparental offspring. These offspring can display combinations of traits that are different from those of either parent. In contrast, offspring that have inherited the same combination of alleles that are found in the chromosomes of their parents are known as nonrecombinant offspring, or parental offspring.



(a) Without crossing over, linked alleles segregate together.

(b) Crossing over can reassort linked alleles.

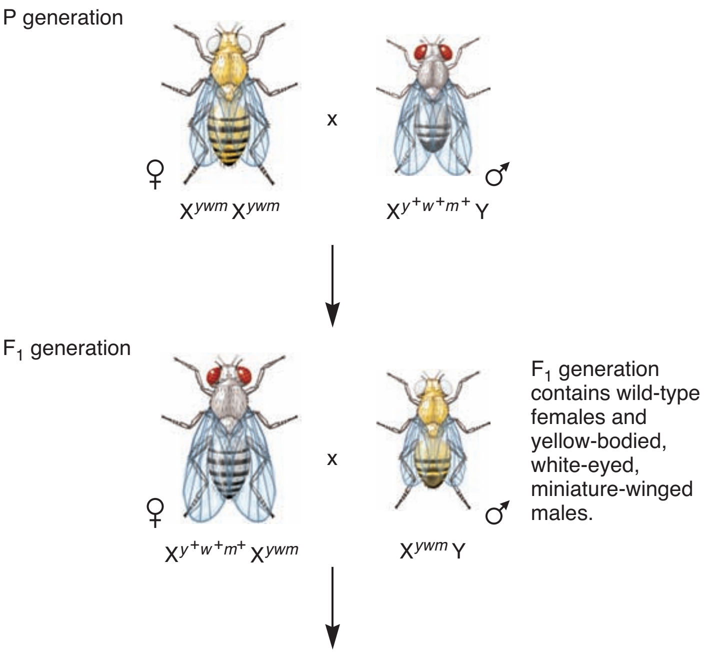
ONLINE ANIMATION

FIGURE 7.2 **Consequences of crossing over during meiosis.** **(a)** In the absence of crossing over, the *A* and *B* alleles and the *a* and *b* alleles are maintained in the same arrangement found in the parental chromosomes. **(b)** Crossing over has occurred in the region between the two genes, producing two recombinant haploid cells with a new combination of alleles.

Concept check: If a crossover began in the short region between gene A and the tip of the chromosome, would this event affect the arrangement of the A and B alleles?

Morgan Provided Evidence for the Linkage of X-Linked Genes and Proposed That Crossing Over Between X Chromosomes Can Occur

The first direct evidence that different genes are physically located on the same chromosome came in 1911 from the studies of Thomas Hunt Morgan, who investigated the inheritance pattern of different traits that had been shown to follow an X-linked pattern of inheritance. Figure 7.3 illustrates an experiment involving three characters that Morgan studied. His parental crosses were wild-type male fruit flies mated to females that had yellow bodies (*yy*), white eyes (*ww*), and miniature wings (*mm*). The wild-type alleles for these three genes are designated y^+ (gray body), w^+ (red eyes), and m^+ (long wings). As expected, the phenotypes of the F_1 generation were wild-type females, and males with yellow bodies, white eyes, and miniature wings. The linkage of these genes was revealed when the F_1 flies were mated to each other and the F_2 generation examined.


P generation
♀
$X^{ywm} X^{ywm}$
x
♂
$X^{y^+w^+m^+} Y$
F_1 generation
♀
$X^{y^+w^+m^+} X^{ywm}$
x
♂
$X^{ywm} Y$
F_1 generation contains wild-type females and yellow-bodied, white-eyed, miniature-winged males.

F_2 generation	Females	Males	Total
Gray body, red eyes, long wings	439	319	758
Gray body, red eyes, miniature wings	208	193	401
Gray body, white eyes, long wings	1	0	1
Gray body, white eyes, miniature wings	5	11	16
Yellow body, red eyes, long wings	7	5	12
Yellow body, red eyes, miniature wings	0	0	0
Yellow body, white eyes, long wings	178	139	317
Yellow body, white eyes, miniature wings	365	335	700

FIGURE 7.3 **Morgan's trihybrid cross involving three X-linked traits in *Drosophila.***

Genes → Traits Three genes that govern body color, eye color, and wing length are all found on the X chromosome. Therefore, the offspring tend to inherit the parental combinations of alleles (y^+ w^+ m^+ or y w m). Figure 7.5 explains how single and double crossovers can create recombinant combinations of alleles.

Concept check: Of the 8 possible phenotypic combinations in the F_2 generation, which ones are the product of a single crossover?

Instead of equal proportions of the eight possible phenotypes, Morgan observed a much higher proportion of the combinations of traits found in the parental generation. He observed 758 flies with gray bodies, red eyes, and long wings, and 700 flies with yellow bodies, white eyes, and miniature wings. The combination of gray body, red eyes, and long wings was found in the males of the parental generation, and the combination of yellow body, white eyes, and miniature wings was the same as the females of the parental generation. Morgan's explanation for this higher proportion of parental combinations was that all three genes are located on the X chromosome and, therefore, tend to be transmitted together as a unit.

However, to fully account for the data shown in Figure 7.3, Morgan needed to explain why a significant proportion of the F_2 generation had a recombinant arrangements of alleles. Along with the two parental phenotypes, five other phenotypic combinations appeared that were not found in the parental generation. How did Morgan explain these data? He considered the studies conducted in 1909 of the Belgian cytologist Frans Alfons Janssens, who observed chiasmata under the microscope and proposed that crossing over involves a physical exchange between homologous chromosomes. Morgan shrewdly realized that crossing over between homologous X chromosomes was consistent with his data. He assumed that crossing over did not occur between the X and Y chromosome and that these three genes are not found on the Y chromosome. With these ideas in mind, he hypothesized that the genes for body color, eye color, and wing length are all located on the same chromosome, namely, the X chromosome. Therefore, the alleles for all three characters are most likely to be inherited together. Due to crossing over, Morgan also proposed that, in the female, the homologous X chromosomes can exchange pieces of chromosomes and create new (recombinant) arrangements of alleles and traits in the F_2 generation.

To appreciate Morgan's proposals, let's simplify his data and consider only two of the three genes, those that affect body color and eye color. If we use the data from Figure 7.3, the following results were obtained:

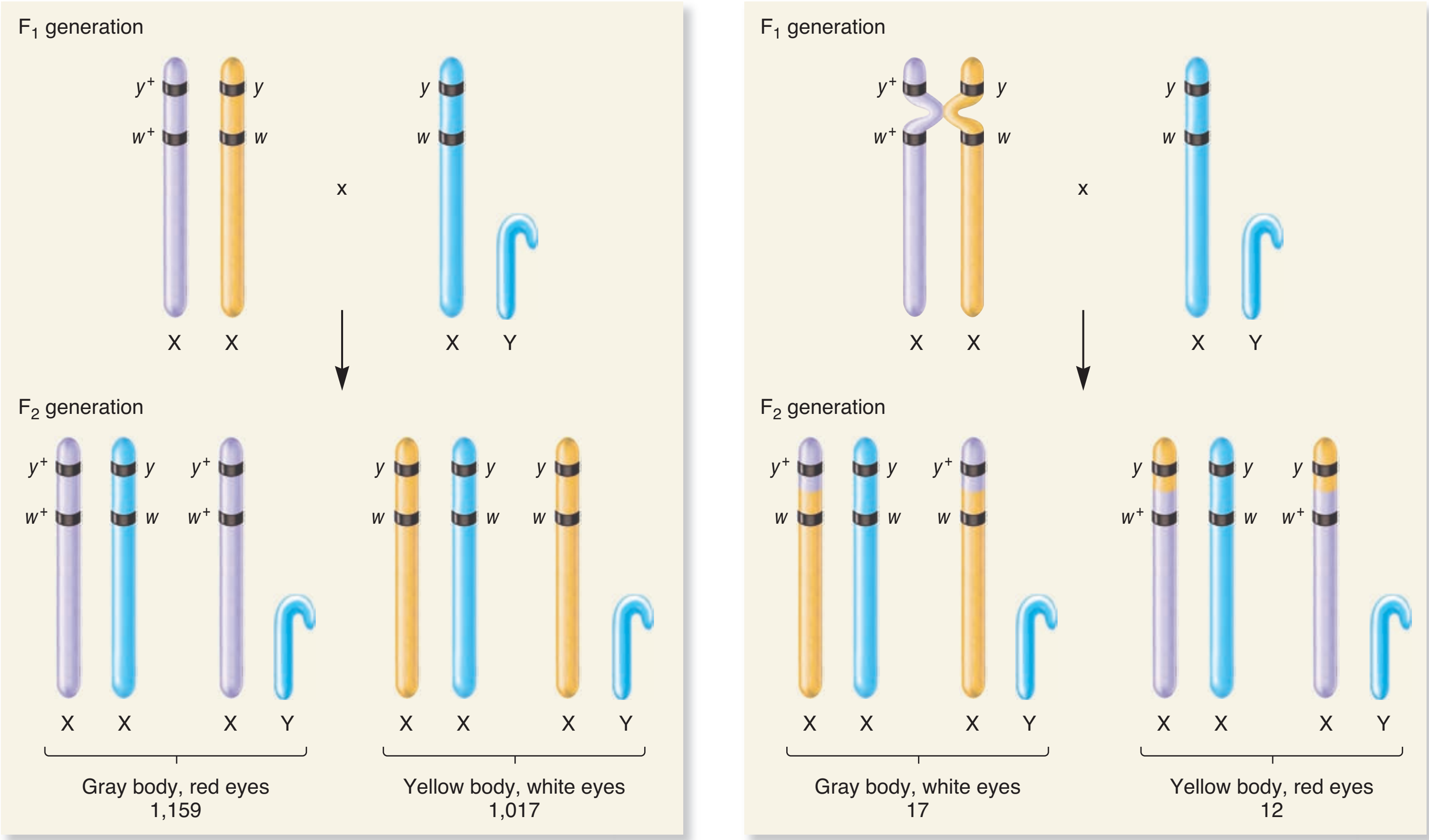
Gray body, red eyes	1159	
Yellow body, white eyes	1017	
Gray body, white eyes	17	Recombinant offspring
Yellow body, red eyes	12	
Total	2205	

Figure 7.4 shows how Morgan's proposals could account for these data. The nonrecombinant offspring with gray bodies and red eyes or yellow body and white eyes were produced when no crossing over had occurred between the two genes (Figure 7.4a). This was the more common situation. By comparison, crossing over could alter the pattern of alleles along each chromosome and account for the recombinant offspring (Figure 7.4b). Why were there relatively few recombinant offspring? These two genes are very close together along the same chromosome, which makes it unlikely that a crossover would be initiated between them. As described next, the distance between two genes is an important factor that determines the relative proportions of recombinant offspring.

The Likelihood of Crossing Over Between Two Genes Depends on the Distance Between Them

For the experiment of Figure 7.3, Morgan also noticed a quantitative difference between the numbers of recombinant offspring when comparing body color and eye color versus eye color and wing length. This quantitative difference is revealed by reorganizing the data of Figure 7.3 by pairs of genes.

Gray body, red eyes	1159	
Yellow body, white eyes	1017	
Gray body, white eyes	17	Recombinant offspring
Yellow body, red eyes	12	
Total	2205	



(a) No crossing over, nonrecombinant offspring

(b) Crossing over, recombinant offspring

FIGURE 7.4 **Morgan's explanation for nonrecombinant and recombinant offspring.** As described in Chapter 2, crossing over actually occurs at the bivalent stage, but for simplicity, this figure shows only two X chromosomes (one of each homolog) rather than four chromatids, which would occur during the bivalent stage of meiosis. Also note that this figure shows only a portion of the X chromosome. A map of the entire X chromosome is shown in Figure 7.7.

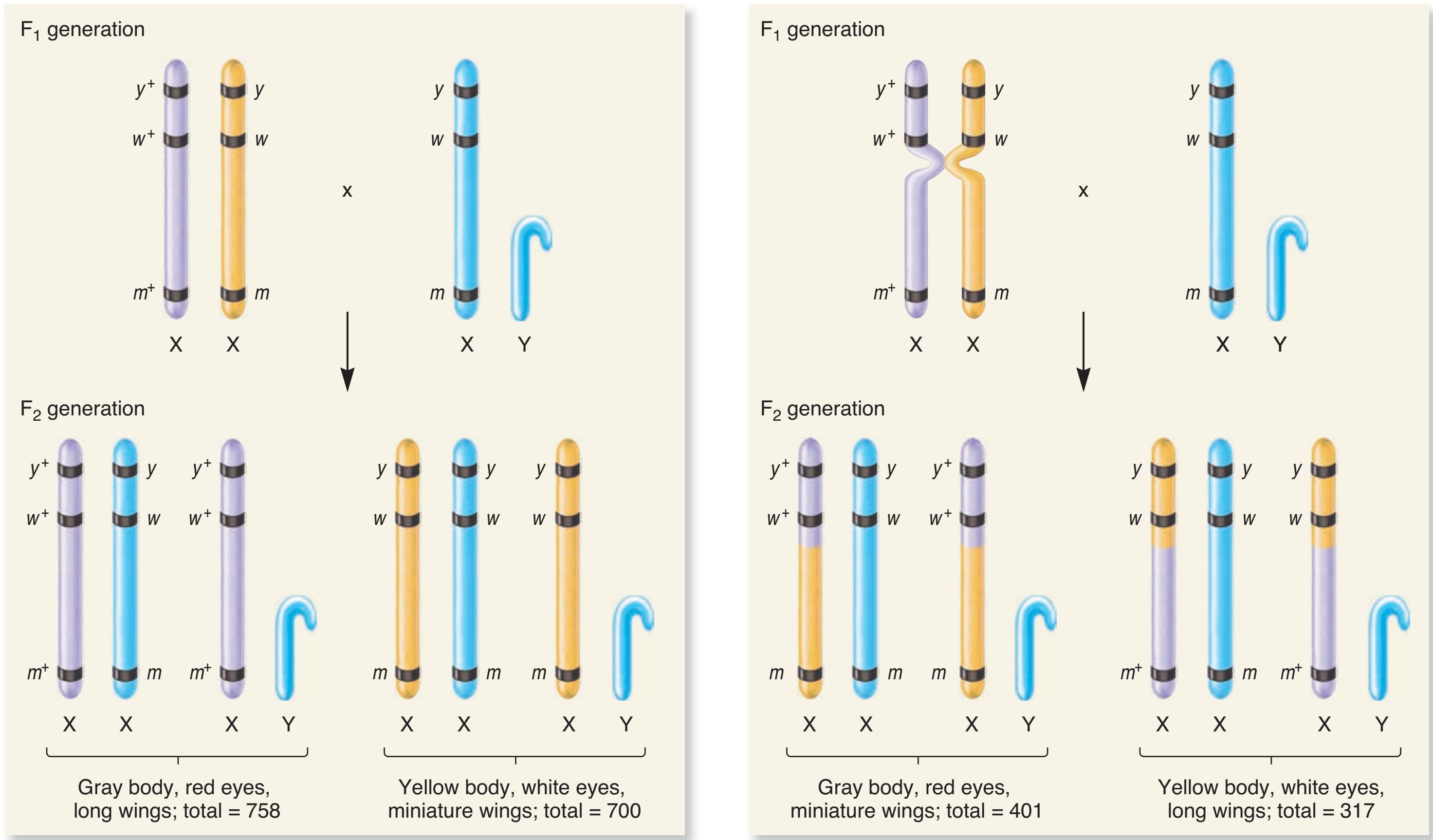
Concept check: *Why are the nonrecombinant offspring more common than the recombinant offspring?*

Red eyes, long wings	770	
White eyes, miniature wings	716	
Red eyes, miniature wings	401	Recombinant offspring
White eyes, long wings	318	
Total	2205	

Morgan found a substantial difference between the numbers of recombinant offspring when pairs of genes were considered separately. Recombinant offspring involving only eye color and wing length were fairly common (401 + 318). In sharp contrast, recombinant offspring for body color and eye color were quite rare (17 + 12).

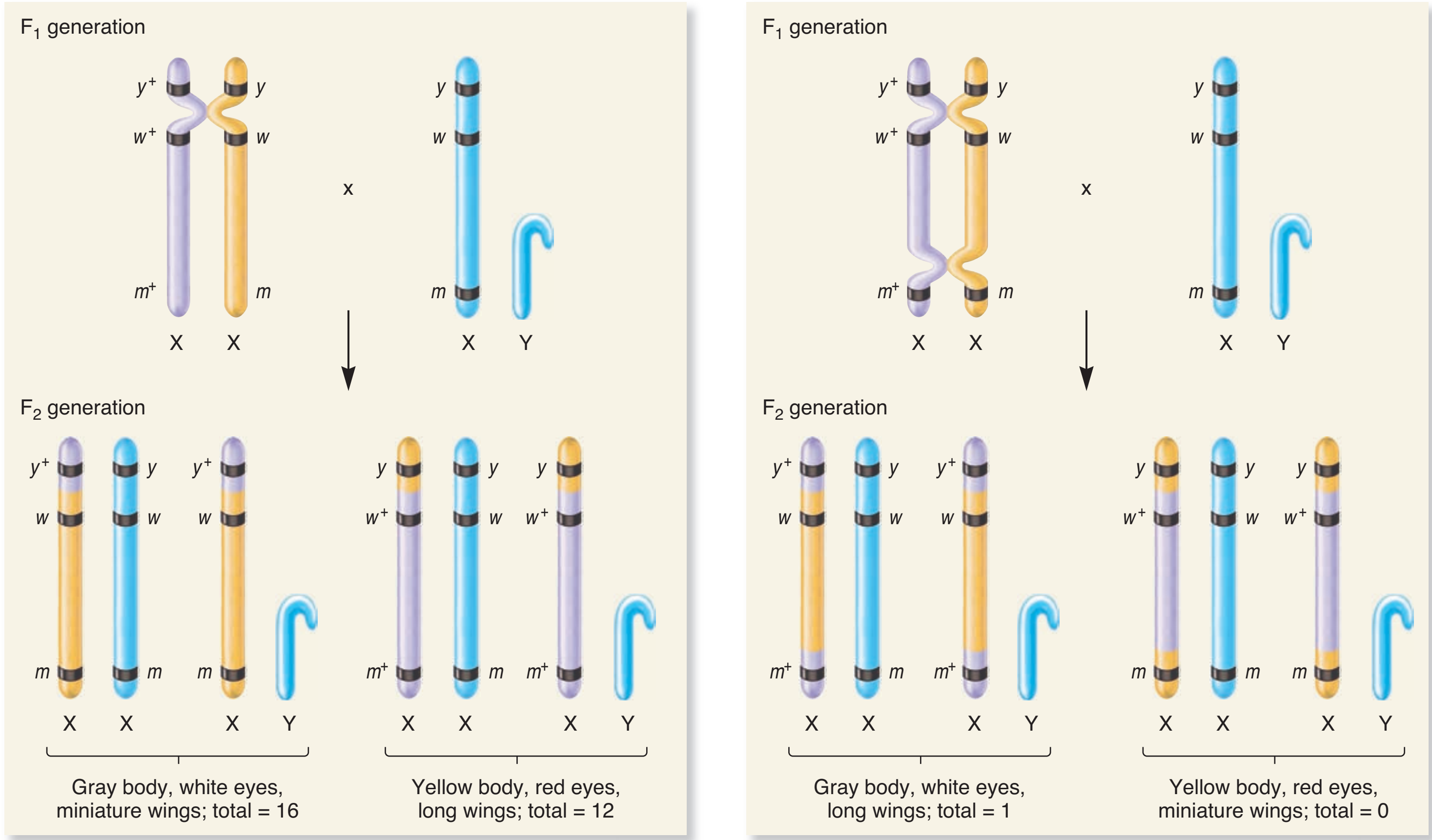
How did Morgan explain these data? He proposed that the likelihood of crossing over depends on the distance between two genes. If two genes are far apart from each other, crossing over is more likely to occur between them than between two genes that are close together.

Figure 7.5 illustrates the possible events that occurred in the F$_1$ female flies of Morgan's experiment. One of the X chromosomes carried all three dominant alleles; the other had all three recessive alleles. During oogenesis in the F$_1$ female flies, crossing over may or may not have occurred in this region of the X chromosome. If no crossing over occurred, the nonrecombinant phenotypes were produced in the F$_2$ offspring (Figure 7.5a). Alternatively, a crossover sometimes occurred between the eye color gene and the wing length gene to produce recombinant offspring with gray bodies, red eyes, and miniature wings or yellow bodies, white eyes, and long wings (Figure 7.5b). According to Morgan's proposal, such an event is fairly likely because these two genes are far apart from each other on the X chromosome. In contrast, he proposed that the body color and eye color genes are very close together, which makes crossing over between them an unlikely event. Nevertheless, it occasionally occurred, yielding offspring with gray bodies, white eyes, and miniature wings, or with yellow bodies, red eyes, and long wings (Figure 7.5c). Finally, it was also possible for two homologous chromosomes to cross over twice (Figure 7.5d). This double crossover is very unlikely. Among the 2205 offspring Morgan examined, he found only 1 fly with a gray body, white eyes, and long wings that could be explained by this phenomenon.


F_1 generation
y^+
w^+
m^+
y
w
m
X X
x
X Y
F_2 generation
X X X Y
Gray body, red eyes, long wings; total = 758
Yellow body, white eyes, miniature wings; total = 700
F_1 generation
F_2 generation
Gray body, red eyes, miniature wings; total = 401
Yellow body, white eyes, long wings; total = 317

(a) No crossing over in this region, very common

(b) Crossover between eye color and wing length genes, fairly common


F_1 generation
y^+
w^+
m^+
y
w
m
X X
x
X Y
F_2 generation
X X X Y
Gray body, white eyes, miniature wings; total = 16
Yellow body, red eyes, long wings; total = 12
F_1 generation
F_2 generation
Gray body, white eyes, long wings; total = 1
Yellow body, red eyes, miniature wings; total = 0

(c) Crosssover between body color and eye color genes, uncommon

(d) Double crossover, very uncommon

FIGURE 7.5 **Morgan's explanation for different proportions of recombinant offspring.** Crossing over is more likely to occur between two genes that are relatively far apart than for two genes that are very close together. A double crossover is particularly uncommon.

Concept check: Why are the types of offspring described in part (b) more numerous than those described in part (c)?

A Chi Square Analysis Can Be Used to Distinguish Between Linkage and Independent Assortment

Now that we have an appreciation for linkage and the production of recombinant offspring, let's consider how an experimenter can objectively decide whether two genes are linked or assort independently. In Chapter 3, we used a chi square analysis to evaluate the goodness of fit between a genetic hypothesis and observed experimental data. This method can similarly be employed to determine if the outcome of a dihybrid cross is consistent with linkage or independent assortment.

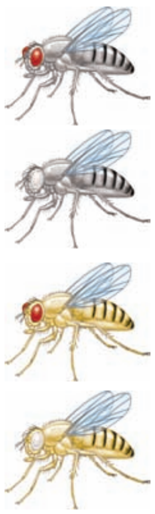
To conduct a chi square analysis, we must first propose a hypothesis. In a dihybrid cross, the standard hypothesis is that the two genes are not linked. This hypothesis is chosen even if the observed data suggest linkage, because an independent assortment hypothesis allows us to calculate the expected number of offspring based on the genotypes of the parents and the law of independent assortment. In contrast, for two linked genes that have not been previously mapped, we cannot calculate the expected number of offspring from a genetic cross because we do not know how likely it is for a crossover to occur between the two genes. Without expected numbers of nonrecombinant and recombinant offspring, we cannot conduct a chi square test. Therefore, we begin with the hypothesis that the genes are not linked. Recall from Chapter 3 that the hypothesis we are testing is called a **null hypothesis,** because it assumes there is no real difference between the observed and expected values. The goal is to determine whether or not the data fit the hypothesis. If the chi square value is low and we cannot reject the null hypothesis, we infer that the genes assort independently. On the other hand, if the chi square value is so high that our hypothesis is rejected, we accept the alternative hypothesis, namely, that the genes are linked.

Of course, a statistical analysis cannot prove that a hypothesis is true. If the chi square value is high, we accept the linkage hypothesis because we are assuming that only two explanations for a genetic outcome are possible: the genes are either linked or not linked. However, if other factors affect the outcome of the cross, such as a decreased viability of particular phenotypes, these may result in large deviations between the observed and expected values and cause us to reject the independent assortment hypothesis even though it may be correct.

As an example of how a chi square analysis can be used to distinguish between linkage and independent assortment, let's reconsider Morgan's data concerning body color and eye color (see Figure 7.4). This cross produced the following offspring: 1159 gray body, red eyes; 1017 yellow body, white eyes; 17 gray body, white eyes; and 12 yellow body, red eyes. However, when a heterozygous female ($X^{y^+w^+}$ X^{yw}) is crossed to a hemizygous male ($X^{yw}Y$), the laws of segregation and independent assortment predict the following outcome:

F_1 male gametes

F_1 female gametes ♀ \ ♂	X^{yw}	Y
$X^{y^+w^+}$	$X^{y^+w^+}X^{yw}$ Gray body, red eyes	$X^{y^+w^+}Y$ Gray body, red eyes
X^{y^+w}	$X^{y^+w}X^{yw}$ Gray body, white eyes	$X^{y^+w}Y$ Gray body, white eyes
X^{yw^+}	$X^{yw^+}X^{yw}$ Yellow body, red eyes	$X^{yw^+}Y$ Yellow body, red eyes
X^{yw}	$X^{yw}X^{yw}$ Yellow body, white eyes	$X^{yw}Y$ Yellow body, white eyes



Mendel's laws predict a 1:1:1:1 ratio among the four phenotypes. The observed data obviously seem to conflict with this expected outcome. Nevertheless, we stick to the strategy just discussed. We begin with the hypothesis that the two genes are not linked and then conduct a chi square analysis to see if the data fit this hypothesis. If the data do not fit, we reject the idea that the genes assort independently and conclude the genes are linked.

A step-by-step outline for applying the chi square test to distinguish between linkage and independent assortment is described next.

Step 1. *Propose a hypothesis.* Even though the observed data appear inconsistent with this hypothesis, we propose that the two genes for eye color and body color obey Mendel's law of independent assortment. This hypothesis allows us to calculate expected values. Because the data seem to conflict with this hypothesis, we actually anticipate that the chi square analysis will allow us to reject the independent assortment hypothesis in favor of a linkage hypothesis. We also assume the alleles follow the law of segregation and the four phenotypes are equally viable.

Step 2. *Based on the hypothesis, calculate the expected values of each of the four phenotypes.* Each phenotype has an equal probability of occurring (see the Punnett square given previously). Therefore, the probability of each phenotype is 1/4. The observed F_2 generation had a total of 2205 individuals. Our next step is to calculate the expected number of offspring with each phenotype when the total equals 2205; 1/4 of the offspring should be each of the four phenotypes:

1/4 × 2205 = 551 (expected number of each phenotype, rounded to the nearest whole number)

Step 3. *Apply the chi square formula, using the data for the observed values (O) and the expected values (E) that have been calculated in step 2.* In this case, the data consist of four phenotypes.

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$\chi^2 = \frac{(1159 - 551)^2}{551} + \frac{(17 - 551)^2}{551} + \frac{(12 - 551)^2}{551} + \frac{(1017 - 551)^2}{551}$$

$$\chi^2 = 670.9 + 517.5 + 527.3 + 394.1 = 2109.8$$

Step 4. *Interpret the calculated chi square value.* This is done with a chi square table, as discussed in Chapter 3. The four phenotypes are based on the law of segregation and the law of independent assortment. By itself, the law of independent assortment predicts only two categories: nonrecombinant and recombinant. Therefore, based on a hypothesis of independent assortment, the degrees of freedom equals $n-1$, which is $2-1$, or 1.

The calculated chi square value, 2109.8, is enormous! This means that the deviation between observed and expected values is very large. With 1 degree of freedom, such a large deviation is expected to occur by chance alone less than 1% of the time (see Table 3.2). Therefore, we reject the hypothesis that the two genes assort independently. As an alternative, we could accept the hypothesis that the genes are linked. Even so, it should be emphasized that rejecting the null hypothesis does not necessarily mean that the linked hypothesis is correct. For example, some of the non-Mendelian inheritance patterns described in Chapter 6 can produce results that do not conform to independent assortment.

Creighton and McClintock Showed That Crossing Over Produced New Combinations of Alleles and Resulted in the Exchange of Segments Between Homologous Chromosomes

As we have seen, Morgan's studies were consistent with the hypothesis that crossing over occurs between homologous chromosomes to produce new combinations of alleles. To obtain direct evidence that crossing over can result in genetic recombination, Harriet Creighton and Barbara McClintock used an interesting strategy involving parallel observations. In studies conducted in 1931, they first made crosses involving two linked genes to produce nonrecombinant and recombinant offspring. Second, they used a microscope to view the structures of the chromosomes in the parents and in the offspring. Because the parental chromosomes had some unusual structural features, they could microscopically distinguish the two homologous chromosomes within a pair. As we will see, this enabled them to correlate the occurrence of recombinant offspring with microscopically observable exchanges in segments of homologous chromosomes.

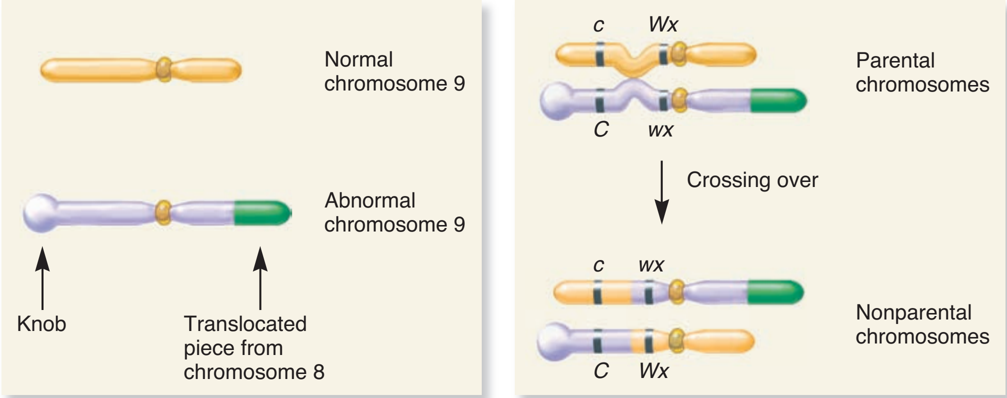
Creighton and McClintock focused much of their attention on the pattern of inheritance of traits in corn. This species has 10 different chromosomes per set, which are named chromosome 1, chromosome 2, chromosome 3, and so on. In previous cytological examinations of corn chromosomes, some strains were found to have an unusual chromosome 9 with a darkly staining knob at one end. In addition, McClintock identified an abnormal version of chromosome 9 that also had an extra piece of chromosome 8 attached at the other end (**Figure 7.6a**). This chromosomal rearrangement is called a translocation.

Creighton and McClintock insightfully realized that this abnormal chromosome could be used to determine if two homologous chromosomes physically exchange segments as a result of crossing over. They knew that a gene was located near the knobbed end of chromosome 9 that provided color to corn kernels. This gene existed in two alleles, the dominant allele *C* (colored) and the recessive allele *c* (colorless). A second gene, located near the translocated piece from chromosome 8, affected the texture of the kernel endosperm. The dominant allele *Wx* caused starchy endosperm, and the recessive *wx* allele caused waxy endosperm. Creighton and McClintock reasoned that a crossover involving a normal chromosome 9 and a knobbed/translocated chromosome 9 would produce a chromosome that had either a knob or a translocation, but not both. These two types of chromosomes would be distinctly different from either of the parental chromosomes (**Figure 7.6b**).

As shown in **Figure 7.6c** (top), if the chromosomes that were transmitted to offspring were not the product of a crossover, the offspring would inherit the colorless and starchy allele or the colored and waxy allele. Alternatively, if crossing over occurred (Figure 7c bottom), the offspring would inherit the colorless and waxy allele or the colored and starchy allele. Creighton and McClintock discovered that offspring with recombinant chromosomes also exhibited a recombinant pair of traits, such as colorless and waxy or colored and starchy. This observation was consistent with idea that the exchange of genetic material due to crossing over was responsible for the pattern of alleles inherited by the recombinant offspring.

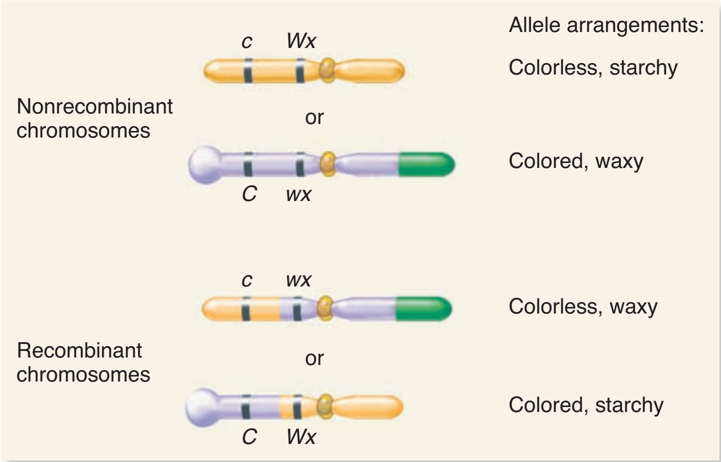
7.2 REVIEWING THE KEY CONCEPTS

- Crossing over can change the combination of alleles along a chromosome and produce recombinant cells and recombinant offspring (see Figure 7.2).
- Morgan discovered genetic linkage in *Drosophila* and proposed that recombinant offspring are produced by crossing over during meiosis (see Figures 7.3, 7.4).
- When genes are linked, the relative proportions of recombinant offspring depends on the distance between the genes (see Figure 7.5).
- A chi square analysis can be used to judge whether or not two genes assort independently.



(a) Normal and abnormal chromosome 9

(b) Crossing over between normal and abnormal chromosome 9



(c) A comparison of allele arrangements on nonrecombinant and recombinant chromosomes

FIGURE 7.6 Crossing over between a normal and abnormal chromosome 9 in corn. **(a)** A normal chromosome 9 in corn is compared with an abnormal chromosome 9 that contains a knob at one end and a translocation at the opposite end. **(b)** The morphological features of the chromosomes differ if crossing over occurs. In this case, a crossover produces a chromosome that contains only a knob at one end and another chromosome that contains only a translocation at the other end. **(c)** The combination of alleles inherited by the offspring also differ if crossing over occurs.

Concept check: In this experiment, what are the two types of characteristics that crossing over can change? Hint: One type is seen only with a microscope, whereas the other type can be seen with the unaided eye.

- Creighton and McClintock were able to correlate the formation of recombinant offspring with the presence of chromosomes that had exchanged pieces due to crossing over (see Figure 7.6).

7.2 COMPREHENSION QUESTIONS

1. With regard to linked genes along the same chromosome, which of the following statements is *false*?
   a. Crossing over is needed to produce nonrecombinant offspring.
   b. Crossing over is needed to produce recombinant offspring.
   c. Crossing over is more likely to separate alleles if they are far apart on the same chromosome.
   d. Crossing over that separates linked alleles occurs during prophase of meiosis I.
2. Morgan observed a higher number of recombinant offspring involving eye color and wing length (401 + 318) than recombinants for body color and eye color (17 + 12). These results occurred because
   a. the genes affecting eye color and wing length are farther apart on the X chromosome than are the genes affecting body color and eye color.
   b. the genes affecting eye color and wing length are closer together on the X chromosome than are the genes affecting body color and eye color.
   c. the gene affecting wing length is not on the X chromosome.
   d. the gene affecting body color is not on the X chromosome.
3. In a chi square analysis involving genes that may be linked, which of the following statements is correct?
   a. An independent assortment hypothesis is not proposed because the data usually suggest linkage.
   b. An independent assortment hypothesis is proposed because it allows you to calculate expected numbers of offspring.
   c. A large chi square value suggests that the observed and expected data are in good agreement.
   d. The hypothesis is rejected when the chi square value is very low.

7.3 GENETIC MAPPING IN PLANTS AND ANIMALS

The purpose of **genetic mapping,** also known as gene mapping, or chromosome mapping, is to determine the linear order and distance of separation among genes that are linked to each other along the same chromosome. Figure 7.7 illustrates a simplified genetic map of *Drosophila melanogaster,* depicting the locations

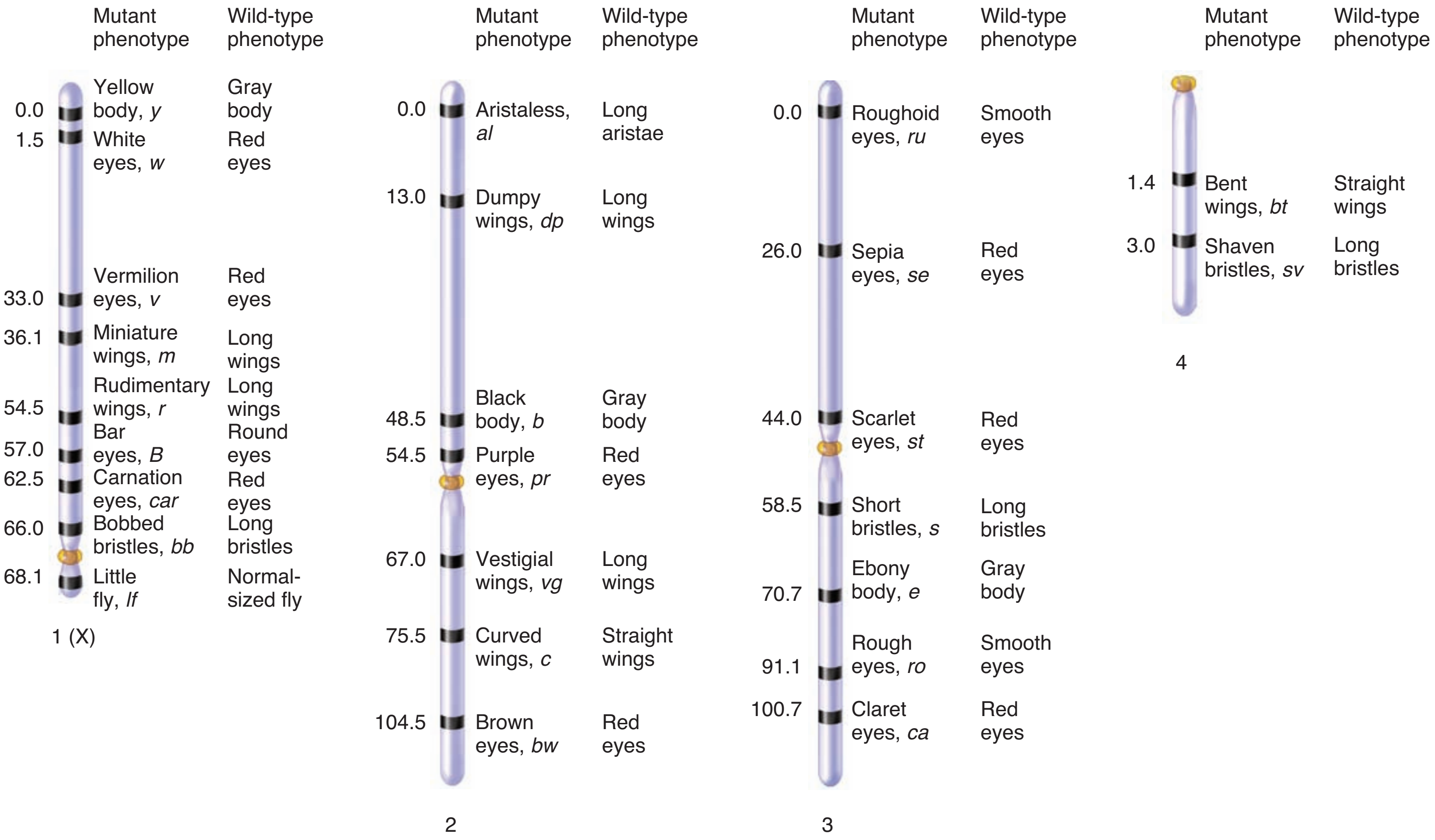

Mutant phenotype
Wild-type phenotype
0.0 Yellow body, y Gray body
1.5 White eyes, w Red eyes
33.0 Vermilion eyes, v Red eyes
36.1 Miniature wings, m Long wings
54.5 Rudimentary wings, r Long wings
57.0 Bar eyes, B Round eyes
62.5 Carnation eyes, car Red eyes
66.0 Bobbed bristles, bb Long bristles
68.1 Little fly, lf Normal-sized fly
1 (X)
0.0 Aristaless, al Long aristae
13.0 Dumpy wings, dp Long wings
48.5 Black body, b Gray body
54.5 Purple eyes, pr Red eyes
67.0 Vestigial wings, vg Long wings
75.5 Curved wings, c Straight wings
104.5 Brown eyes, bw Red eyes
2
0.0 Roughoid eyes, ru Smooth eyes
26.0 Sepia eyes, se Red eyes
44.0 Scarlet eyes, st Red eyes
58.5 Short bristles, s Long bristles
70.7 Ebony body, e Gray body
91.1 Rough eyes, ro Smooth eyes
100.7 Claret eyes, ca Red eyes
3
1.4 Bent wings, bt Straight wings
3.0 Shaven bristles, sv Long bristles
4

FIGURE 7.7 **A simplified genetic linkage map of *Drosophila melanogaster*.** This simplified map illustrates a few of the many thousands of genes that have been identified in this organism.

Concept check: *List five reasons why genetic maps are useful.*

of many different genes along the individual chromosomes. As shown here, each gene has its own unique **locus—**the site where the gene is found within a particular chromosome. For example, the gene designated *brown eyes* (*bw*), which affects eye color, is located near one end of chromosome 2. The gene designated *black body* (*b*), which affects body color, is found near the middle of the same chromosome.

Why is genetic mapping useful? First, it allows geneticists to understand the overall complexity and genetic organization of a particular species. The genetic map of a species portrays the underlying basis for the inherited traits that an organism displays. In some cases, the known locus of a gene within a genetic map can help molecular geneticists to clone that gene and thereby obtain greater information about its molecular features. In addition, genetic maps are useful from an evolutionary point of view. A comparison of the genetic maps for different species can improve our understanding of the evolutionary relationships among those species.

Along with these scientific uses, genetic maps have many practical benefits. For example, many human genes that play a role in human disease have been genetically mapped. This information can be used to diagnose and perhaps someday treat inherited human diseases. It can also help genetic counselors predict the likelihood that a couple will produce children with certain inherited diseases. In addition, genetic maps are gaining increasing importance in agriculture. A genetic map can provide plant and animal breeders with helpful information for improving agriculturally important strains through selective breeding programs.

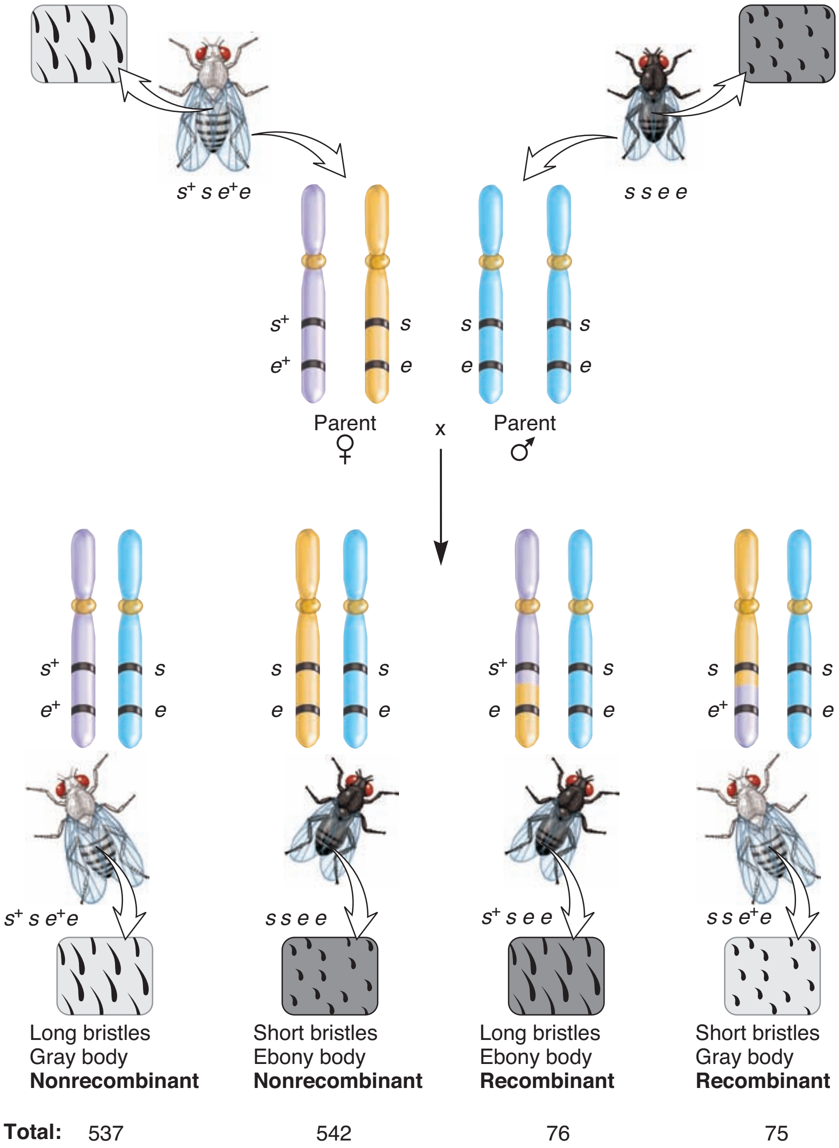
In this section, we will examine traditional genetic mapping techniques that involve an analysis of crosses of individuals that are heterozygous for two or more genes. The frequency of recombinant offspring due to crossing over provides a way to deduce the linear order of genes along a chromosome. As depicted in Figure 7.7, this linear arrangement of genes is known as a **genetic linkage map.** This approach has been useful for analyzing organisms that are easily crossed and produce a large number of offspring in a short period of time. Genetic linkage maps have been constructed for several plant species and certain species of animals, such as *Drosophila*. For many organisms, however, traditional mapping approaches are difficult due to long generation times or the inability to carry out experimental crosses (as in humans). Fortunately, many alternative methods of gene mapping have been developed to replace the need to carry out crosses. As described in Chapter 21, molecular approaches are increasingly used to map genes.

The Frequency of Recombination Between Two Genes Is Used to Compute Map Distance Along a Chromosome

Genetic mapping allows us to estimate the relative distances between linked genes based on the likelihood that a crossover will occur between them. If two genes are very close together on the same chromosome, a crossover is unlikely to begin in the region between them. However, if two genes are very far apart, a crossover is more likely to be initiated in this region and thereby recombine the alleles of the two genes. Experimentally, the basis for genetic mapping is that the percentage of recombinant offspring is correlated with the distance between two genes. If two genes are far apart, many recombinant offspring will be produced. However, if two genes are close together, very few recombinant offspring will be observed.

To interpret a genetic mapping experiment, the experimenter must know if the characteristics of an offspring are due to crossing over during meiosis in a parent. This is accomplished by conducting a **testcross.** Most testcrosses are between an individual that is heterozygous for two or more genes and an individual that is recessive and homozygous for the same genes. The goal of the testcross is to determine if recombination has occurred during meiosis in the heterozygous parent. Therefore, genetic mapping is based on the level of recombination that occurs in just one parent—the heterozygote. In a testcross, new combinations of alleles cannot occur in the gametes of the other parent, which is homozygous for these genes.

Figure 7.8 illustrates how a testcross provides an experimental strategy to distinguish between nonrecombinant and recombinant offspring. This cross concerns two linked genes affecting bristle length and body color in fruit flies. The dominant



INTERACTIVE EXERCISE

FIGURE 7.8 Use of a testcross to distinguish between nonrecombinant and recombinant offspring. The cross involves one *Drosophila* parent that is homozygous recessive for short bristles (*ss*) and ebony body (*ee*), and one parent heterozygous for both genes ($s^+s\ e^+e$). (*Note: Drosophila* geneticists normally designate the short allele as *ss* and a homozygous fly with short bristles as *ssss*. In this case, the allele causing short bristles is designated with a single *s* to avoid confusion between the allele designation and the genotype of the fly. Also, crossing over does not occur during sperm formation in *Drosophila,* which is unusual among eukaryotes. Therefore, the heterozygote in a testcross involving *Drosophila* must be the female.)

Concept check: When and in which fly or flies did crossing over occur in order to produce the recombinant offspring?

(wild-type) alleles are s^+ (long bristles) and e^+ (gray body) and the recessive alleles are *s* (short bristles) and *e* (ebony body). One parent displays both recessive traits. Therefore, we know this parent is homozygous for the recessive alleles of the two genes (*ss ee*). The other parent is heterozygous for the linked genes affecting bristle length and body color. This parent was produced from a cross involving a true-breeding wild-type fly and a true-breeding fly with short bristles and an ebony body. Therefore, in this heterozygous parent, we know that the s^+ and e^+ alleles are located on one chromosome and the corresponding *s* and *e* alleles are located on the homologous chromosome. In this case, when the wild-type alleles are all located on the same chromosome and the corresponding mutant alleles are found on the homolog, this is called **coupling** or ***cis* configuration.** Alternatively, if s^+ and *e* were on one chromosome and *s* and e^+ were on the homolog, this would be called **repulsion** or ***trans* configuration.** The *cis* or *trans* configuration determines which phenotypes are most common among the offspring of a testcross.

Let's take a look at the four possible types of offspring these parents can produce (see bottom of Figure 7.8). The offspring's phenotypes are long bristles, gray body; short bristles, ebony body; long bristles, ebony body; and short bristles, gray body. All four types of offspring have inherited a chromosome carrying the *s* and *e* alleles from their homozygous parent (shown on the right in each pair). Focus your attention on the other chromosome. The offspring with long bristles and a gray body have inherited a chromosome carrying the s^+ and e^+ alleles from the heterozygous parent. This chromosome is not the product of a crossover. The offspring with short bristles and an ebony body have inherited a chromosome carrying the *s* and *e* alleles from the heterozygous parent. Again, this chromosome is not the product of a crossover.

The other two types of offspring, however, can be produced only if crossing over has occurred in the region between these two genes. Those with long bristles and an ebony body or short bristles and a gray body have inherited a chromosome that is the product of a crossover during meiosis in the heterozygous parent. As noted in Figure 7.8, the recombinant offspring are fewer in number than are the nonrecombinant offspring.

The frequency of recombination can be used as an estimate of the physical distance between two genes on the same chromosome. The **map distance** is defined as the number of recombinant offspring divided by the total number of offspring, multiplied by 100. We can calculate the map distance between these two genes using this formula:

$$\begin{aligned}\text{Map distance} &= \frac{\text{Number of recombinant offspring}}{\text{Total number of offspring}} \times 100\\ &= \frac{76 + 75}{537 + 542 + 76 + 75} \times 100\\ &= 12.3 \text{ map units}\end{aligned}$$

The units of distance are called **map units (mu),** or sometimes **centiMorgans (cM)** in honor of Thomas Hunt Morgan. One map unit is equivalent to a 1% frequency of recombination. In this example, we conclude that the *s* and *e* alleles are 12.3 mu apart from each other along the same chromosome.

Alfred Sturtevant Used the Frequency of Crossing Over in Dihybrid Crosses to Produce the First Genetic Map

In 1913, the first individual to construct a (very small) genetic map was Alfred Sturtevant, an undergraduate who spent time in the laboratory of Thomas Hunt Morgan. Sturtevant wrote: "In conversation with Morgan . . . I suddenly realized that the variations in the strength of linkage, already attributed by Morgan to differences in the spatial separation of the genes, offered the possibility of determining sequences [of different genes] in the linear dimension of a chromosome. I went home and spent most of the night (to the neglect of my undergraduate homework) in producing the first chromosome map, which included the sex-linked genes, *y*, *w*, *v*, *m*, and *r*, in the order and approximately the relative spacing that they still appear on the standard maps."

In the experiment of **Figure 7.9**, Sturtevant considered the outcome of crosses involving six different mutant alleles that altered the phenotype of flies. All of these alleles were known to be recessive and X-linked. They are *y* (yellow body color), *w* (white eye color), w^e (eosin eye color), *v* (vermilion eye color), *m* (miniature wings), and *r* (rudimentary wings). The *w* and w^e alleles are alleles of the same gene. In contrast, the *v* allele (vermilion eye color) is an allele of a different gene that also affects eye color. The two alleles that affect wing length, *m* and *r*, are also in different genes. Therefore, Sturtevant studied the inheritance of six recessive alleles, but since *w* and w^e are alleles of the same gene, his genetic map contained only five genes. The corresponding wild-type alleles are y^+ (gray body), w^+ (red eyes), v^+ (red eyes), m^+ (long wings), and r^+ (long wings).

▶ THE HYPOTHESIS

When genes are located on the same chromosome, the distance between the genes can be estimated from the proportion of recombinant offspring. This provides a way to map the order of genes along a chromosome.

TESTING THE HYPOTHESIS — FIGURE 7.9 The first genetic mapping experiment.

Concept check: *Describe the basic strategy that is followed in this experiment.*

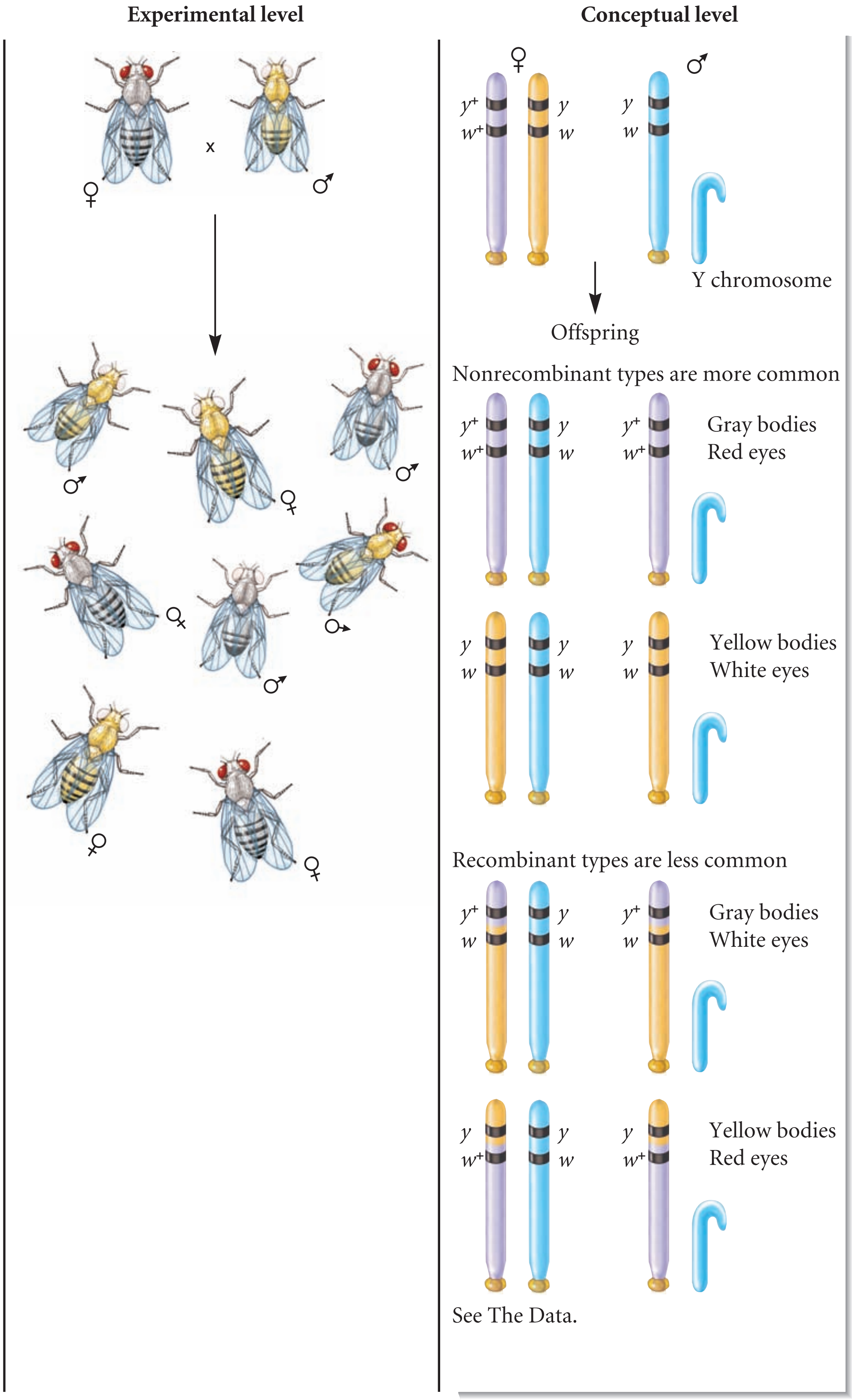
Starting materials: Sturtevant began with several different strains of *Drosophila* that contained the six alleles already described.

1. Cross a female that is heterozygous for two different genes to a male that is hemizygous recessive for the same two genes. In this example, cross a female that is $X^{y^+w^+}X^{yw}$ to a male that is $X^{yw}Y$.

   This strategy was employed for many dihybrid combinations of the six alleles already described.

2. Observe the outcome of the crosses.

3. Calculate the percentages of offspring that are the result of crossing over (number of recombinant/total).



Experimental level
Conceptual level
♀
x
♂
y^+
w^+
y
w
Y chromosome
Offspring
Nonrecombinant types are more common
Gray bodies
Red eyes
Yellow bodies
White eyes
Recombinant types are less common
Gray bodies
White eyes
Yellow bodies
Red eyes
See The Data.

THE DATA

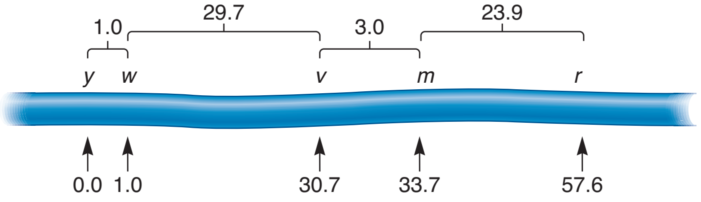
Alleles Concerned	*Number Recombinant/Total Number*	*Percent Recombinant Offspring*
y and w/w^e	214/21,736	1.0
y and *v*	1464/4551	32.2
y and *m*	115/324	35.5
y and *r*	260/693	37.5
w/w^e and *v*	471/1584	29.7
w/w^e and *m*	2062/6116	33.7
w/w^e and *r*	406/898	45.2
v and *m*	17/573	3.0
v and *r*	109/405	26.9

Data from Alfred H. Sturtevant (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool 14*, 43–59.

INTERPRETING THE DATA

As shown in Figure 7.9, Sturtevant made pairwise testcrosses and then counted the number of offspring in the four phenotypic categories. Two of the categories were nonrecombinant and two were recombinant, requiring a crossover between the X chromosomes in the female heterozygote. Let's begin by contrasting the results between particular pairs of genes, shown in the data. In some dihybrid crosses, the percentage of recombinant offspring was rather low. For example, dihybrid crosses involving the *y* allele and the *w* or w^e allele yielded 1% recombinant offspring. This result suggested that these two genes are very close together. By comparison, other dihybrid crosses showed a higher percentage of recombinant offspring. For example, crosses involving the *v* and *r* alleles produced 26.9% recombinant offspring. These two genes are expected to be farther apart.

To construct his map, Sturtevant began with the assumption that the map distances would be more accurate between genes that are closely linked. Therefore, his map is based on the distance between *y* and *w* (1.0), *w* and *v* (29.7), *v* and *m* (3.0), and *v* and *r* (26.9). He also considered other features of the data to deduce the order of the genes. For example, the percentage of crossovers between *w* and *m* was 33.7. The percentage of crossovers between *w* and *v* was 29.7, suggesting that *v* is between *w* and *m*, but closer to *m*. The proximity of *v* and *m* is confirmed by the low percentage of crossovers between *v* and *m* (3.0). Sturtevant collectively considered the data and proposed the genetic map shown here:



In this genetic map, Sturtevant began at the *y* allele and mapped the genes from left to right. For example, the *y* and *v* alleles are 30.7 mu apart, and the *v* and *m* alleles are 3.0 mu apart. This study by Sturtevant was a major breakthrough, because it showed how to map the locations of genes along chromosomes by making the appropriate crosses.

If you look carefully at Sturtevant's data, you will notice a few observations that do not agree very well with his genetic map. For example, the percentage of recombinant offspring for the *y* and *r* dihybrid cross was 37.5 (but the map distance is 57.6), and the crossover percentage between *w* and *r* was 45.2 (but the map distance is 56.6). As the percentage of recombinant offspring approaches a value of 50%, this value becomes a progressively more inaccurate measure of actual map distance. To correct for this inaccuracy, geneticists have developed **mapping functions,** which describe the quantitative relationship between recombination frequencies and physical distances along chromosomes (**Figure 7.10**). What is the basis for this inaccuracy? When the distance between two genes is large, the likelihood of multiple crossovers in the region between them causes the observed number of recombinant offspring to underestimate the physical distance.

Multiple crossovers set a quantitative limit on the relationship between map distance and the percentage of recombinant offspring. Even though two different genes can be on the same chromosome and more than 50 mu apart, a testcross is expected to yield a maximum of only 50% recombinant offspring. What accounts for this 50% limit? The answer lies in the pattern of multiple crossovers. A single crossover in the region between two genes produces only 50% recombinant chromosomes (see Figure

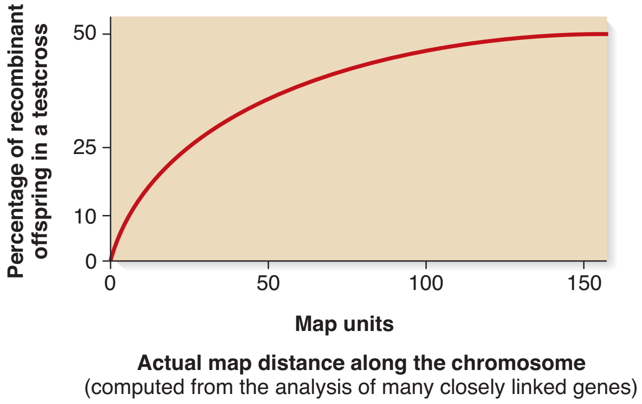


FIGURE 7.10 **Relationship between the percentage of recombinant offspring in a testcross and the actual map distance between genes.** The *y*-axis depicts the percentage of recombinant offspring that would be observed in a dihybrid testcross. The actual map distance, shown on the *x*-axis, is calculated by analyzing the percentages of recombinant offspring from a series of many dihybrid crosses involving closely linked genes. Even though two genes may be more than 50 mu apart, the percentage of recombinant offspring will not exceed 50%.

Concept check: *What phenomenon explains why the maximum percentage of recombinant offspring does not exceed 50%?*

7.2b). Therefore, to exceed a 50% recombinant level, it seems necessary to have multiple crossovers within a tetrad. However, let's consider double crossovers. As shown in the figure to Solved Problem S5 at the end of the chapter, a double crossover between two genes could involve four, three, or two chromatids, which would yield 100%, 50%, or 0% recombinants, respectively. Because all of these double crossovers are equally likely, we take the average of them to determine the maximum recombination frequency. This average equals 50%. Therefore, when two different genes are more than 50 mu apart, they follow the law of independent assortment in a testcross and only 50% recombinants are observed.

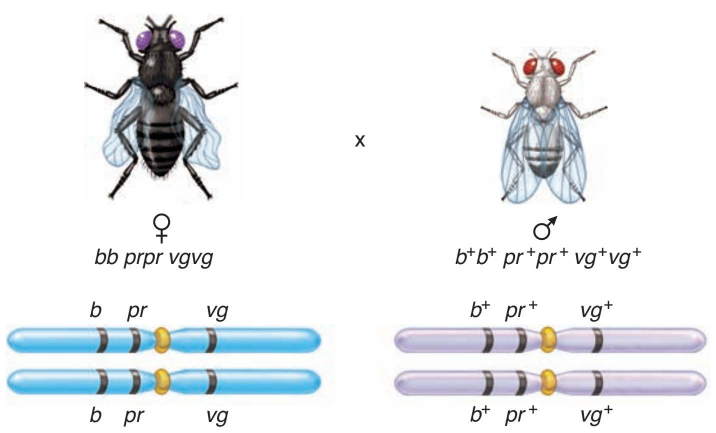
A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

Trihybrid Crosses Can Be Used to Determine the Order and Distance Between Linked Genes

Thus far, we have considered the construction of genetic maps using dihybrid testcrosses to compute map distance. The data from trihybrid crosses can yield additional information about map distance and gene order. In a trihybrid cross, the experimenter crosses two individuals that differ in three characters. The following experiment outlines a common strategy for using trihybrid crosses to map genes. In this experiment, the parental generation consists of fruit flies that differ in body color, eye color, and wing shape. We must begin with true-breeding lines so we know which alleles are initially linked to each other on the same chromosome. In this example, all of the dominant alleles are linked on the same chromosome.

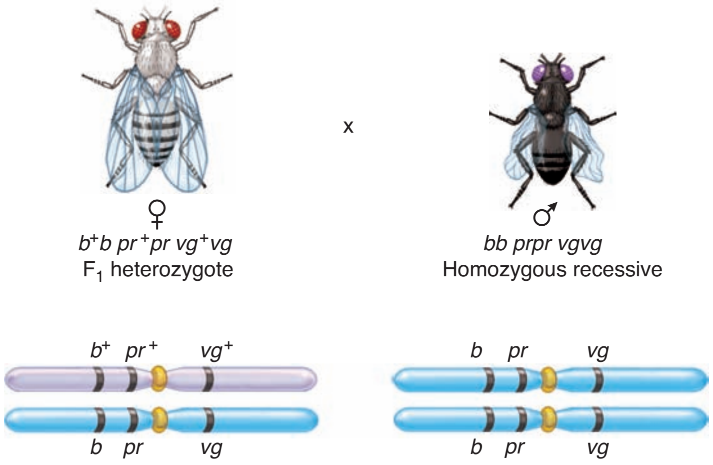
Step 1. *Cross two true-breeding strains that differ with regard to three alleles.* In this example, we cross a fly that has a black body (*bb*), purple eyes (*prpr*), and vestigial wings (*vgvg*) to a homozygous wild-type fly with a gray body (b^+b^+), red eyes (pr^+pr^+), and long wings (vg^+vg^+):

Parental flies



The goal in this step is to obtain F_1 individuals that are heterozygous for all three genes. In the F_1 heterozygotes, all dominant alleles are located on one chromosome, and all recessive alleles are on the other homologous chromosome.

Step 2. *Perform a testcross by mating F_1 female heterozygotes to male flies that are homozygous recessive for all three alleles (bb prpr vgvg).*



During gametogenesis in the heterozygous female F_1 flies, crossovers may produce new combinations of the three alleles.

Step 3. *Collect data for the F_2 generation.* As shown in **Table 7.1**, eight phenotypic combinations are possible. An analysis of the F_2 generation flies allows us to map these three genes. Because the three genes exist as two alleles each, we have 2^3, or 8, possible combinations of offspring. If these alleles assorted independently, all eight combinations would occur in equal proportions. However, we see from the data that the proportions of the eight phenotypes are far from equal.

The genotypes of the parental generation correspond to the phenotypes gray body, red eyes, and long wings, and black body, purple eyes, and vestigial wings. In crosses involving linked genes, the nonrecombinant phenotypes occur most frequently in the offspring. The remaining six phenotypes are due to crossing over.

The double crossover is always expected to be the least frequent category of offspring. Two of the phenotypes—gray body, purple eyes, and long wings; and black body, red eyes, and vestigial wings—arose from a double crossover between two pairs of genes. Also, the combination of traits in the double crossover tells us which gene is in the middle. When a chromatid undergoes a double crossover, the gene in the middle becomes separated from the other two genes at either end.

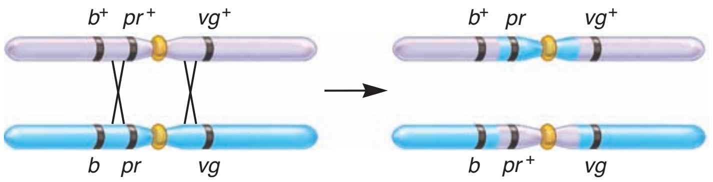
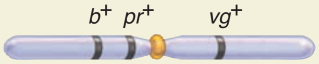
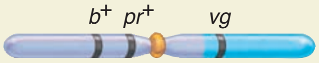
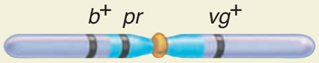
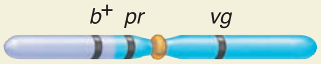
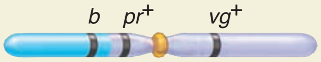
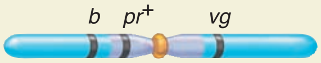
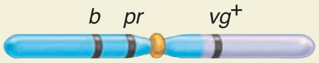
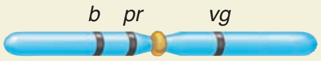


TABLE 7.1

Data from a Trihybrid Cross (see Steps 2 and 3)

Phenotype	Number of Observed Offspring (males and females)	Chromosome Inherited from F_1 Female
Gray body, red eyes, long wings	411	 b^+ pr^+ vg^+
Gray body, red eyes, vestigial wings	61	 b^+ pr^+ vg
Gray body, purple eyes, long wings	2	 b^+ pr vg^+
Gray body, purple eyes, vestigial wings	30	 b^+ pr vg
Black body, red eyes, long wings	28	 b pr^+ vg^+
Black body, red eyes, vestigial wings	1	 b pr^+ vg
Black body, purple eyes, long wings	60	 b pr vg^+
Black body, purple eyes, vestigial wings	412	 b pr vg
Total	1005	

In the double-crossover categories, the recessive purple eye allele is separated from the other two recessive alleles. When mated to a homozygous recessive fly in the testcross, this yields flies with gray bodies, purple eyes, and long wings; or flies with black bodies, red eyes, and vestigial wings. This observation indicates that the gene for eye color lies between the genes for body color and wing shape.

Step 4. *Calculate the map distance between pairs of genes.* To do this, we need to understand which gene combinations are nonrecombinant and which are recombinant. The recombinant offspring are due to crossing over in the heterozygous female parent. If you look back at step 2, you can see the arrangement of alleles in the heterozygous female parent in the absence of crossing over. Let's consider this arrangement with regard to gene pairs:

b^+ is linked to pr^+, and b is linked to pr
pr^+ is linked to vg^+, and pr is linked to vg
b^+ is linked to vg^+, and b is linked to vg

With regard to body color and eye color, the recombinant offspring have gray bodies and purple eyes (2 + 30) or black bodies and red eyes (28 + 1). As shown along the right side of Table 7.1, these offspring were produced by crossovers in the female parents. The total number of these recombinant offspring is 61. The map distance between the body color and eye color genes is

$$\text{Map distance} = \frac{61}{944 + 61} \times 100 = 6.1 \text{ mu}$$

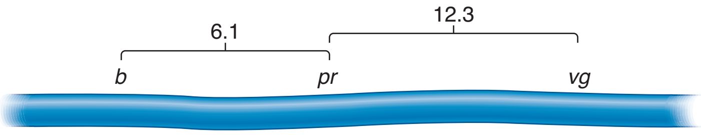
With regard to eye color and wing shape, the recombinant offspring have red eyes and vestigial wings (61 + 1) or purple eyes and long wings (2 + 60). The total number is 124. The map distance between the eye color and wing shape genes is

$$\text{Map distance} = \frac{124}{881 + 124} \times 100 = 12.3 \text{ mu}$$

With regard to body color and wing shape, the recombinant offspring have gray bodies and vestigial wings (61 + 30) or black bodies and long wings (28 + 60). The total number is 179. The map distance between the body color and wing shape genes is

$$\text{Map distance} = \frac{179}{826 + 179} \times 100 = 17.8 \text{ mu}$$

Step 5. *Construct the map.* Based on the map unit calculation, the body color (b) and wing shape (vg) genes are farthest apart. The eye color gene (pr) must lie in the middle. As mentioned earlier, this order of genes is also confirmed by the pattern of traits found in the double crossovers. To construct the map, we use the distances between the genes that are closest together.



In our example, we have placed the body color gene first and the wing shape gene last. The data also are consistent with a map in which the wing shape gene comes first and the body color gene comes last. In detailed genetic maps, the locations of genes are mapped relative to the centromere.

You may have noticed that our calculations underestimate the distance between the body color and wing shape genes. We obtained a value of 17.8 mu even though the distance seems to be 18.4 mu when we add together the distance between body color and eye color genes (6.1 mu) and the distance between eye color and wing shape genes (12.3 mu). What accounts for this discrepancy? The answer is double crossovers. If you look at the data in Table 7.1, the offspring with gray bodies, purple eyes, and long wings or those with black bodies, red eyes, and vestigial wings are due to a double crossover. From a phenotypic perspective, these offspring are not recombinant with regard to the body color and wing shape alleles. Even so, we know that they arose from a double crossover between these two genes. Therefore, we should consider these crossovers when calculating the distance between the body color and wing shape genes. In this case, three offspring (2 + 1) were due to double crossovers. Because they are double crossovers, we multiply 2 times the number of double crossovers (2 + 1) and add this number to our previous value of recombinant offspring:

$$\text{Map distance} = \frac{179 + 2(2 + 1)}{826 + 179} \times 100 = 18.4 \text{ mu}$$

Interference Can Influence the Number of Double Crossovers That Occur in a Short Region

In Chapter 3, we considered the product rule to determine the probability that two independent events will both occur. The product rule allows us to predict the expected likelihood of a double crossover, provided we know the individual probabilities of each single crossover. Let's reconsider the data of the trihybrid testcross just described to see if the frequency of double crossovers is what we would expect based on the product rule. If each crossover is an independent event, we can multiply the likelihood of a single crossover between *b* and *pr* (0.061) times the likelihood of a single crossover between *pr* and *vg* (0.123). The product rule predicts

Expected likelihood of a double crossover $= 0.061 \times 0.123 = 0.0075 = 0.75\%$

Expected number of offspring due to a double crossover, based on a total of 1005 offspring produced $= 1005 \times 0.0075 = 7.5$

In other words, we would expect about 7 or 8 offspring to be produced as a result of a double crossover. The observed number of offspring was only 3 (namely, 2 with gray bodies, purple eyes, and long wings, and 1 with a black body, red eyes, and vestigial wings). What accounts for the lower number? This lower-than-expected value is probably not due to random sampling error. Instead, the likely cause is a common genetic phenomenon known as **positive interference,** in which the occurrence of a crossover in one region of a chromosome decreases the probability that a second crossover will occur nearby. In other words, the first crossover interferes with the ability to form a second crossover in the immediate vicinity. To provide interference with a quantitative value, we first calculate the coefficient of coincidence (C), which is the ratio of the observed number of double crossovers to the expected number.

$$C = \frac{\text{Observed number of double crossovers}}{\text{Expected number of double crossovers}}$$

Interference (I) is expressed as

$$I = 1 - C$$

For the data of the trihybrid testcross, the observed number of crossovers is 3 and the expected number is 7.5, so the coefficient of coincidence equals $3/7.5 = 0.40$. In other words, only 40% of the expected number of double crossovers were actually observed. The value for interference equals $1 - 0.4 = 0.60$, or 60%. This means that 60% of the expected number of crossovers did not occur.

7.3 REVIEWING THE KEY CONCEPTS

- A genetic linkage map is a diagram that portrays the order and relative spacing of genes along one or more chromosomes (see Figure 7.7).
- A testcross can be performed to map the distance between two or more genes (see Figure 7.8).
- Sturtevant was the first scientist to conduct testcrosses and map the order of a few genes along the X chromosome in *Drosophila* (see Figure 7.9).
- Due to the effects of multiple crossovers, the map distance between two genes obtained from a testcross cannot exceed 50% (see Figure 7.10).
- The data from a trihybrid cross can be used to map genes (see Table 7.1).
- Positive interference refers to the phenomenon that the number of double crossovers in a given region is less than expected based on the frequency of single crossovers.

7.3 COMPREHENSION QUESTIONS

Answer the multiple choice questions based on the following experiments:

Parental generation: True-breeding flies with red eyes and long wings were crossed to flies with white eyes and miniature wings. All F_1 offspring had red eyes and long wings.

The F_1 female flies were then crossed to males with white eyes and miniature wings. The following results were obtained for the F_2 generation:

129 red eyes, long wings

133 white eyes, miniature wings

71 red eyes, miniature wings

67 white eyes, long wings

1. What is/are the phenotypes of the recombinant offspring of the F_2 generation?
   a. Red eyes, long wings
   b. White eyes, miniature wings
   c. Red eyes, long wings and white eyes, miniature wings
   d. Red eyes, miniature wings and white eyes, long wings
2. The recombinant offspring of the F_2 generation were produced by crossing over that occurred
   a. during spermatogenesis in the parental generation males.
   b. during oogenesis in the parental generation females.
   c. during spermatogenesis in the F_1 males.
   d. during oogenesis in the F_1 females.
3. What is the map distance between these two genes?
   a. 32.3 mu
   b. 34.5 mu
   c. 16.2 mu
   d. 17.3 mu

7.4 MITOTIC RECOMBINATION

Thus far, we have considered how the arrangement of linked alleles along a chromosome can be rearranged by crossing over. This event can produce cells and offspring with a recombinant set of traits. In these cases, crossing over occurred during meiosis, when the homologous chromosomes replicate and form bivalents.

In multicellular organisms, the union of egg and sperm is followed by many cellular divisions, which occur in conjunction with mitotic divisions of the cell nuclei. As discussed in Chapter 2, mitosis normally does not involve the homologous pairing of chromosomes to form a bivalent. Therefore, crossing over during mitosis is expected to occur much less frequently than during meiosis. Nevertheless, it does happen on rare occasions. Mitotic crossing over may produce a pair of recombinant chromosomes that have a new combination of alleles, an event known as **mitotic recombination.** If it occurs during an early stage of embryonic development, the daughter cells containing one recombinant and one nonrecombinant chromosome continue to divide many times to produce a patch of tissue in the adult. This may result in a portion of tissue with characteristics different from those of the rest of the organism.

In 1936, Curt Stern identified unusual patches on the bodies of certain *Drosophila* strains. He was working with strains carrying X-linked alleles affecting body color and bristle morphology (**Figure 7.11**). A recessive allele confers yellow body color (*y*), and another recessive allele causes shorter body bristles that look

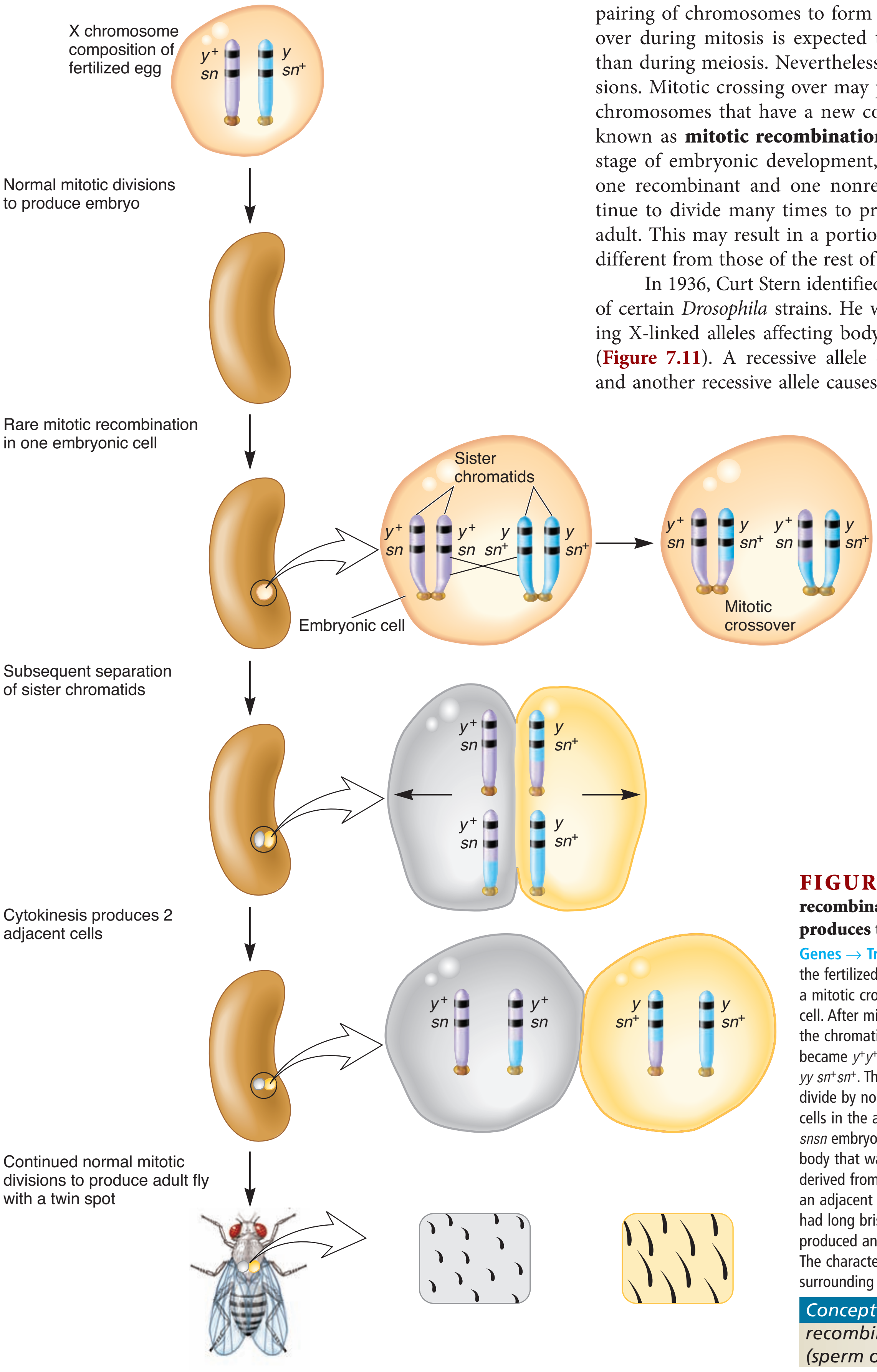

X chromosome composition of fertilized egg
y^+ sn
y sn^+
Normal mitotic divisions to produce embryo
Rare mitotic recombination in one embryonic cell
Sister chromatids
y^+ sn
y^+ sn
y sn^+
y sn^+
Embryonic cell
y^+ sn
y sn^+
y^+ sn
y sn^+
Mitotic crossover
Subsequent separation of sister chromatids
y^+ sn
y sn^+
y^+ sn
y sn^+
Cytokinesis produces 2 adjacent cells
y^+ sn
y^+ sn
y sn^+
y sn^+
Continued normal mitotic divisions to produce adult fly with a twin spot

FIGURE 7.11 Mitotic recombination in *Drosophila* that produces twin spotting.

Genes → Traits In this illustration, the genotype of the fertilized egg was $y^+y\ sn^+sn$. During development, a mitotic crossover occurred in a single embryonic cell. After mitotic crossing over, the separation of the chromatids occurred, so one embryonic cell became $y^+y^+\ snsn$ and the adjacent sister cell became $yy\ sn^+sn^+$. The embryonic cells then continued to divide by normal mitosis to produce an adult fly. The cells in the adult that were derived from the $y^+y^+\ snsn$ embryonic cell produced a spot on the adult body that was gray and had singed bristles. The cells derived from the $yy\ sn^+sn^+$ embryonic cell produced an adjacent spot on the body that was yellow and had long bristles. In this case, mitotic recombination produced an unusual trait known as a twin spot. The characteristics of this twin spot differ from the surrounding tissue, which is gray with long bristles.

Concept check: *Does mitotic recombination occur in a gamete (sperm or egg cell) or in a somatic cell?*

singed (*sn*). The corresponding wild-type alleles result in gray body color (y^+) and long bristles (sn^+). Females that are $y^+y\ sn^+sn$ are expected to have gray body color and long bristles. This was generally the case. However, when Stern carefully observed the bodies of these female flies under a low-power microscope, he occasionally noticed places in which two adjacent regions were different from the rest of the body—a twin spot. He concluded that twin spotting was too frequent to be explained by the random positioning of two independent single spots that happened to occur close together. How then did Stern explain the phenomenon of twin spotting? He proposed that twin spots are due to a single mitotic recombination within one cell during embryonic development.

As shown in Figure 7.11, the X chromosomes of the fertilized egg are $y^+\ sn$ and $y\ sn^+$. During development, a rare crossover can occur during mitosis to produce two adjacent daughter cells that are $y^+y^+\ snsn$ and $yy\ sn^+sn^+$. As embryonic development proceeds, the cell on the left continues to divide to produce many cells, eventually producing a patch on the body that has gray color with singed bristles. The daughter cell next to it produces a patch of yellow body color with long bristles. These two adjacent patches—a twin spot—are surrounded by cells that are $y^+y\ sn^+sn$ and have gray color and long bristles. Twin spots provide evidence that mitotic recombination occasionally occurs.

7.4 REVIEWING THE KEY CONCEPTS

- Mitotic recombination can occur rarely and may lead to twin spots (see Figure 7.11).

7.4 COMPREHENSION QUESTIONS

1. The process of mitotic recombination involves the
    a. exchange of chromosomal regions between homologs during gamete formation.
    b. exchange of chromosomal regions between homologs during the division of somatic cells.
    c. reassortment of alleles that occurs at fertilization.
    d. reassortment of alleles that occurs during gamete formation.

KEY TERMS

Page 129. genetic map
Page 130. synteny, genetic linkage, linkage, linkage groups, dihybrid cross, trihybrid cross
Page 131. crossing over, bivalent, nonrecombinant, parental, genetic recombination, recombinant, nonparental
Page 135. null hypothesis
Page 137. genetic mapping
Page 138. locus, genetic linkage map
Page 139. testcross
Page 140. coupling, *cis* configuration, repulsion, *trans* configuration, map distance, map units (mu), centiMorgans (cM)
Page 142. mapping functions
Page 145. positive interference
Page 146. mitotic recombination

CHAPTER SUMMARY

7.1 Overview of Linkage

- Linkage has two related meanings. First, it refers to genes that are located on the same chromosome. Second, linkage also means that the alleles of two or more genes tend to be transmitted as a unit because they are relatively close on the same chromosome.
- Bateson and Punnett discovered the first example of genetic linkage in sweet peas (see Figure 7.1).

7.2 Relationship Between Linkage and Crossing Over

- Crossing over can change the combination of alleles along a chromosome and produce recombinant cells and recombinant offspring (see Figure 7.2).
- Morgan discovered genetic linkage in *Drosophila* and proposed that recombinant offspring are produced by crossing over during meiosis (see Figures 7.3, 7.4).
- When genes are linked, the relative proportions of recombinant offspring depends on the distance between the genes (see Figure 7.5).
- A chi square analysis can be used to judge whether or not two genes assort independently.
- Creighton and McClintock were able to correlate the formation of recombinant offspring with the presence of chromosomes that had exchanged pieces due to crossing over (see Figure 7.6).

7.3 Genetic Mapping in Plants and Animals

- A genetic linkage map is a diagram that portrays the order and relative spacing of genes along one or more chromosomes (see Figure 7.7).
- A testcross can be performed to map the distance between two or more genes (see Figure 7.8).
- Sturtevant was the first scientist to conduct testcrosses and map the order of a few genes along the X chromosome in *Drosophila* (see Figure 7.9).
- Due to the effects of multiple crossovers, the map distance between two genes obtained from a testcross cannot exceed 50% (see Figure 7.10).
- The data from a trihybrid cross can be used to map genes (see Table 7.1).

- Positive interference refers to the phenomenon that the number of double crossovers in a given region is less than expected based on the frequency of single crossovers.

7.4 Mitotic Recombination

- Mitotic recombination can occur rarely and may lead to twin spots (see Figure 7.11).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. In the garden pea, orange pods (*orp*) are recessive to green pods (*Orp*), and sensitivity to pea mosaic virus (*mo*) is recessive to resistance to the virus (*Mo*). A plant with orange pods and sensitivity to the virus was crossed to a true-breeding plant with green pods and resistance to the virus. The F_1 plants were then testcrossed to plants with orange pods and sensitivity to the virus. The following results were obtained:

160 orange pods, virus-sensitive
165 green pods, virus-resistant
36 orange pods, virus-resistant
39 green pods, virus-sensitive
400 total

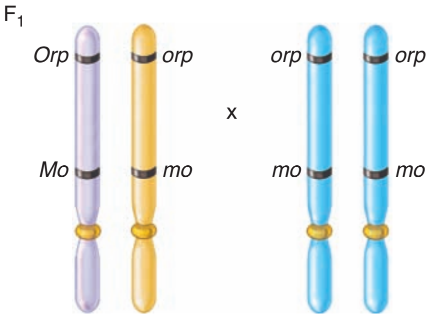
A. Conduct a chi square analysis to see if these genes are linked.

B. If they are linked, calculate the map distance between the two genes.

Answer:

A. Chi square analysis.

1. Our hypothesis is that the genes are not linked.
2. Calculate the predicted number of offspring based on the hypothesis. The testcross is:



The predicted outcome of this cross under our hypothesis is a 1:1:1:1 ratio of plants with the four possible phenotypes. In other words, 1/4 should have the phenotype orange pods, virus-sensitive; 1/4 should have green pods, virus-resistant; 1/4 should have orange pods, virus-resistant; and 1/4 should have green pods, virus-sensitive. Because a total of 400 offspring were produced, our hypothesis predicts 100 offspring in each category.

3. Calculate the chi square.

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$\chi^2 = \frac{(160 - 100)^2}{100} + \frac{(165 - 100)^2}{100} + \frac{(36 - 100)^2}{100} + \frac{(39 - 100)^2}{100}$$

$$\chi^2 = 36 + 42.3 + 41 + 37.2 = 156.5$$

4. Interpret the chi square value. The calculated chi square value is quite large. This indicates that the deviation between observed and expected values is very high. For 1 degree of freedom in Table 3.2, such a large deviation is expected to occur by chance alone less than 1% of the time. Therefore, we reject the hypothesis that the genes assort independently. As an alternative, we may infer that the two genes are linked.

B. Calculate the map distance.

$$\text{Map distance} = \frac{(\text{Number of recombinant offspring})}{\text{Total number of offspring}} \times 100$$

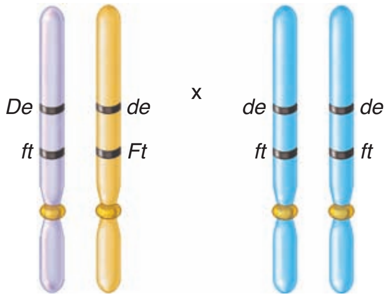
$$= \frac{36 + 39}{36 + 39 + 160 + 165} \times 100$$

$$= 18.8 \text{ mu}$$

The genes are approximately 18.8 mu apart.

S2. Two recessive disorders in mice—droopy ears and flaky tail—are caused by genes that are located 6 mu apart on chromosome 3. A true-breeding mouse with normal ears (*De*) and a flaky tail (*ft*) was crossed to a true-breeding mouse with droopy ears (*de*) and a normal tail (*Ft*). The F_1 offspring were then crossed to mice with droopy ears and flaky tails. If this testcross produced 100 offspring, what is the expected outcome?

Answer: The testcross is:



The nonrecombinant offspring are

Dede ftft	Normal ears, flaky tail
dede Ftft	Droopy ears, normal tail

The recombinant offspring are

dede ftft	Droopy ears, flaky tail
Dede Ftft	Normal ears, normal tail

Because the two genes are located 6 mu apart on the same chromosome, 6% of the offspring will be recombinants. Therefore, the expected outcome for 100 offspring is

3 droopy ears, flaky tail
3 normal ears, normal tail
47 normal ears, flaky tail
47 droopy ears, normal tail

S3. The following X-linked recessive traits are found in fruit flies: vermilion eyes are recessive to red eyes, miniature wings are recessive to long wings, and sable body is recessive to gray body. A cross was made between wild-type males with red eyes, long wings, and gray bodies to females with vermilion eyes, miniature wings, and sable bodies. The heterozygous females from this cross, which had red eyes, long wings, and gray bodies, were then crossed to males with vermilion eyes, miniature wings, and sable bodies. The following outcome was obtained:

Males and Females
1320 vermilion eyes, miniature wings, sable body
1346 red eyes, long wings, gray body
102 vermilion eyes, miniature wings, gray body
90 red eyes, long wings, sable body
42 vermilion eyes, long wings, gray body
48 red eyes, miniature wings, sable body
2 vermilion eyes, long wings, sable body
1 red eyes, miniature wings, gray body

Calculate the map distance between the three genes.

Answer:

The first step is to determine the order of the three genes. We can do this by evaluating the pattern of inheritance in the double crossovers. The double-crossovers occur with the lowest frequency. Thus, the double crossovers are vermilion eyes, long wings, and sable body, and red eyes, miniature wings, and gray body. Compared with the nonrecombinant pattern of alleles (vermilion eyes, miniature wings, sable body and red eyes, long wings, gray body), the gene for wing length has been reassorted. Two flies have long wings associated with vermilion eyes and sable body, and one fly has miniature wings associated with red eyes and gray body. Taken together, these results indicate that the wing length gene is found in between the eye color and body color genes.

Eye color——wing length——body color

We now calculate the distance between eye color and wing length, and between wing length and body color. To do this, we consider the data according to gene pairs:

vermilion eyes, miniature wings = 1320 + 102 = 1422
red eyes, long wings = 1346 + 90 = 1436
vermilion eyes, long wings = 42 + 2 = 44
red eyes, miniature wings = 48 + 1 = 49

The recombinants are vermilion eyes, long wings and red eyes, miniature wings. The map distance between these two genes is

$$(44 + 49)/(1422 + 1{,}436 + 44 + 49) \times 100 = 3.2 \text{ mu}$$

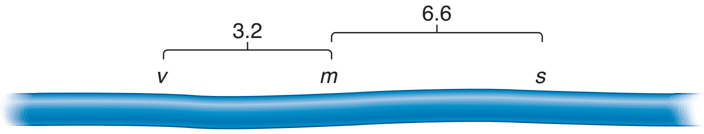
Likewise, the other gene pair is wing length and body color.

miniature wings, sable body = 1320 + 48 = 1368
long wings, gray body = 1346 + 42 = 1388
miniature wings, gray body = 102 + 1 = 103
long wings, sable body = 90 + 2 = 92

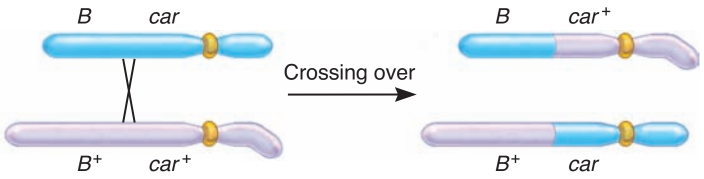
The recombinants are miniature wings, gray body and long wings, sable body. The map distance between these two genes is

$$(103 + 92)/(1368 + 1388 + 103 + 92) \times 100 = 6.6 \text{ mu}$$

With these data, we can produce the following genetic map:



S4. Around the same time as the study of Creighton and McClintock, described in Figure 7.6, Curt Stern conducted similar experiments with *Drosophila*. He had strains of flies with microscopically detectable abnormalities in the X chromosome. In one case, the X chromosome was shorter than normal due to a deletion at one end. In another case, the X chromosome was longer than normal because an extra piece of the Y chromosome was attached at the other end of the X chromosome, where the centromere is located. He had female flies that had both abnormal chromosomes. On the short X chromosome, a recessive allele (*car*) was located that results in carnation-colored eyes, and a dominant allele (*B*) that causes bar-shaped eyes was also found on this chromosome. On the long X chromosome were located the wild-type alleles for these two genes (designated car^+ and B^+), which confer red eyes and round eyes, respectively. Stern realized that a crossover between the two X chromosomes in such female flies would result in recombinant chromosomes that would be cytologically distinguishable from the parental chromosomes. If a crossover occurred between the *B* and *car* genes on the X chromosome, this is expected to produce a normal-sized X chromosome and an abnormal chromosome with a deletion at one end and an extra piece of the Y chromosome at the other end.



Stern crossed these female flies to male flies that had a normal-length X chromosome with the *car* allele and the allele for round eyes (*car* B^+). Using a microscope, he could discriminate between the morphologies of parental chromosomes—like those contained within the original parental flies—and recombinant chromosomes that may be found in the offspring. What would be the predicted phenotypes and chromosome characteristics in the offspring if crossing over did or did not occur between the X chromosomes in the female flies of this cross?

Answer: To demonstrate that genetic recombination is due to crossing over, Stern needed to correlate recombinant phenotypes (due to genetic recombination) with the inheritance of recombinant chromosomes (due to crossing over). Because he knew the arrangement of alleles in the female flies, he could predict the phenotypes of nonrecombinant and recombinant offspring. The male flies could contribute the *car* and B^+ alleles (on a cytologically normal X chromosome) or contribute a Y chromosome. In the absence of crossing over, the female flies could contribute a short X chromosome with the *car* and *B* alleles or a long X chromosome with the car^+ and B^+ alleles. If crossing over occurred in the region between these two genes, the female flies would contribute recombinant X chromosomes. One possible recombinant X chromosome would be normal-sized and carry the *car* and B^+ alleles, and the other recombinant X chromosome would be deleted at one end with a piece of the Y chromosome at the other end and carry the car^+ and *B* alleles. When combined with an X or Y chromosome from the males, the nonrecombinant offspring would have carnation, bar eyes or wild-type eyes; the recombinant offspring would have carnation, round eyes or red, bar eyes.

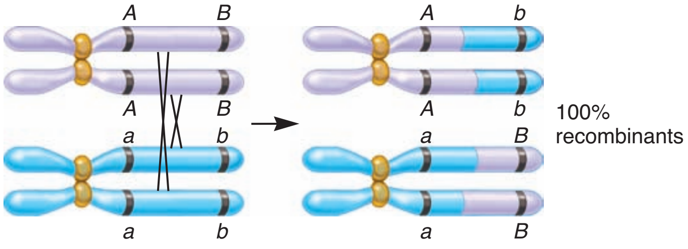
	Male gametes			
♀ Female gametes	♂ $carB^+$	Y	**Phenotype**	**X chromosome from female**
$carB$	$carB$ $carB^+$	$carB$ Y	Carnation, bar eyes	Short X chromosome
car^+B^+	car^+B^+ $carB^+$	car^+B^+ Y	Red, round eyes	Long X chromosome with a piece of Y
$carB^+$	$carB^+$ $carB^+$	$carB^+$ Y	Carnation, round eyes	Normal-sized X chromosome
car^+B	car^+B $carB^+$	car^+B Y	Red, bar eyes	Short X chromosome with a piece of Y

The results shown in the Punnett square are the actual results that Stern observed. His interpretation was that crossing over between homologous chromosomes—in this case, the X chromosome—accounts for the formation of offspring with recombinant phenotypes.

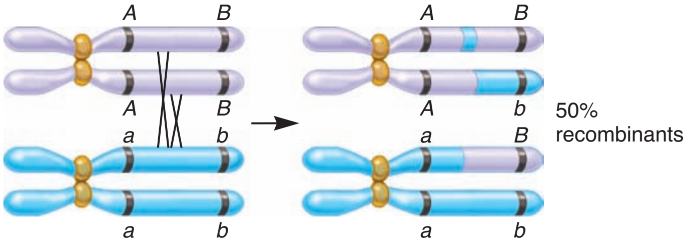
S5. A limit exists in the relationship between map distance and the percentage of recombinant offspring. Even though two genes on the same chromosome may be much more than 50 mu apart, we do not expect to obtain greater than 50% recombinant offspring in a testcross. You may be wondering why this is so. The answer lies in the pattern of multiple crossovers. At the pachytene stage of meiosis, a single crossover in the region between two genes produces only 50% recombinant chromosomes (see Figure 7.2b). Therefore, to exceed a 50% recombinant level, it would seem necessary to have multiple crossovers within the tetrad.

Let's suppose that two genes are far apart on the same chromosome. A testcross is made between a heterozygous individual, *AaBb*, and a homozygous individual, *aabb*. In the heterozygous individual, the dominant alleles (*A* and *B*) are linked on the same chromosome, and the recessive alleles (*a* and *b*) are linked on the same chromosome. Draw out all of the possible double crossovers (between two, three, or four chromatids) and determine the average number of recombinant offspring, assuming an equal probability of all of the double-crossover possibilities.

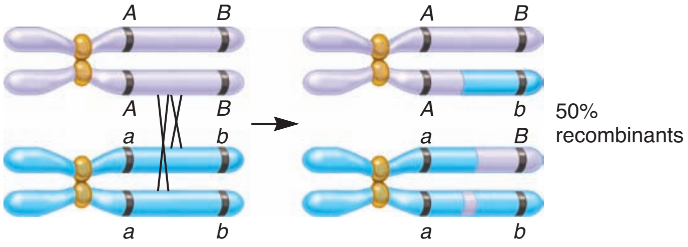
Answer: A double crossover between the two genes could involve two chromatids, three chromatids, or four chromatids. The possibilities for all types of double crossovers are shown here:



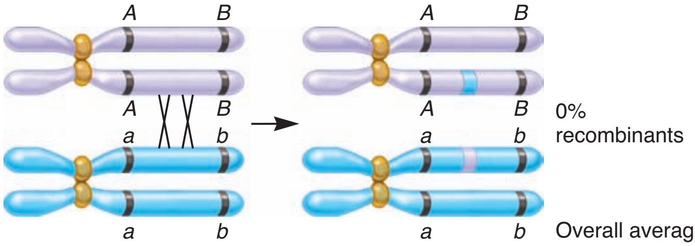
Double crossover (involving 4 chromatids)



Double crossover (involving 3 chromatids)



Double crossover (involving 3 chromatids)



Overall average is 50% for all 4 possibilities.

Double crossover (involving 2 chromatids)

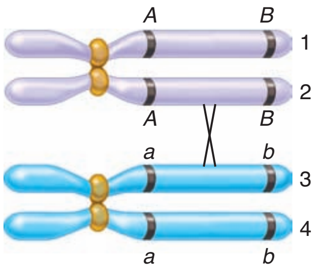
This drawing considers the situation where two crossovers are expected to occur in the region between the two genes. Because the tetrad is composed of two pairs of homologs, a double crossover between homologs could occur in several possible ways. In this illustration, the crossover on the right has occurred first. Because all of these double crossing over events are equally probable, we take the average of them to determine the maximum recombination frequency. This average equals 50%.

Conceptual Questions

C1. What is the difference in meaning between the terms *genetic recombination* and *crossing over*?

C2. When applying a chi square approach in a linkage problem, explain why an independent assortment hypothesis is used.

C3. A crossover has occurred in the bivalent shown here:



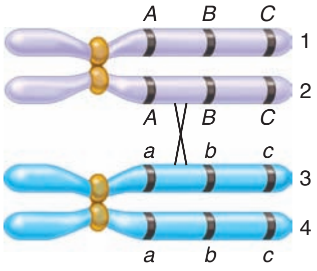
If a second crossover occurs in the same region between these two genes, which two chromatids would be involved to produce the following outcomes?

A. 100% recombinants

B. 0% recombinants

C. 50% recombinants

C4. A crossover has occurred in the bivalent shown here:



What is the outcome of this single-crossover event? If a second crossover occurs somewhere between *A* and *C*, explain which two chromatids it would involve and where it occurs (i.e., between which two genes) to produce the types of chromosomes shown here:

A. *A B C*, *A b C*, *a B c*, and *a b c*

B. *A b c*, *A b c*, *a B C*, and *a B C*

C. *A B c*, *A b c*, *a B C*, and *a b C*

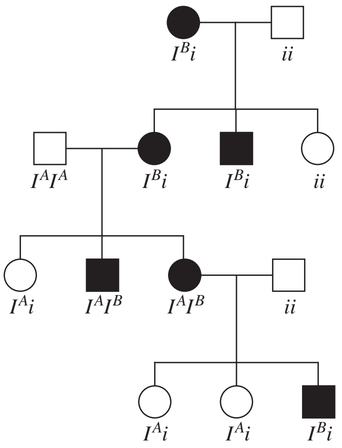
D. *A B C*, *A B C*, *a b c*, and *a b c*

C5. A diploid organism has a total of 14 chromosomes and about 20,000 genes per haploid genome. Approximately how many genes are in each linkage group?

C6. If you try to throw a basketball into a basket, the likelihood of succeeding depends on the size of the basket. It is more likely that you will get the ball into the basket if the basket is bigger. In your own words, explain how this analogy also applies to the idea that the likelihood of crossing over is greater when two genes are far apart than when they are close together.

C7. By conducting testcrosses, researchers have found that the sweet pea has seven linkage groups. How many chromosomes would you expect to find in leaf cells?

C8. In humans, a rare dominant disorder known as nail-patella syndrome causes abnormalities in the fingernails, toenails, and kneecaps. Researchers have examined family pedigrees with regard to this disorder and, within the same pedigree, also examined the individuals with regard to their blood types. (A description of blood genotypes is found in Chapter 5.) In the following pedigree, individuals affected with nail-patella disorder are shown with filled symbols. The genotype of each individual with regard to their ABO blood type is also shown. Does this pedigree suggest any linkage between the gene that causes nail-patella syndrome and the gene that causes blood type?



C9. When true-breeding mice with brown fur and short tails (*BBtt*) were crossed to true-breeding mice with white fur and long tails (*bbTT*), all F_1 offspring had brown fur and long tails. The F_1 offspring were crossed to mice with white fur and short tails. What are the possible phenotypes of the F_2 offspring? Which F_2 offspring are recombinant, and which are nonrecombinant? What are the ratios of the F_2 offspring if independent assortment is taking place? How are the ratios affected by linkage?

C10. Though we often think of genes in terms of the phenotypes they produce (e.g., curly leaves, flaky tail, brown eyes), the molecular function of most genes is to encode proteins. Many cellular proteins function as enzymes. The table that follows describes the map distances between six different genes that encode six different enzymes: *Ada*, adenosine deaminase; *Hao-1*, hydroxyacid oxidase-1; *Hdc*, histidine decarboxylase; *Odc-2*, ornithine decarboxylase-2; *Sdh-1*, sorbitol dehydrogenase-1; and *Ass-1*, arginosuccinate synthetase-1.

Map distances between two genes:

	Ada	*Hao-1*	*Hdc*	*Odc-2*	*Sdh-1*	*Ass-1*
Ada		*14*		*8*	*28*	
Hao-1	*14*		*9*		*14*	
Hdc		*9*		*15*	*5*	
Odc-2	*8*		*15*			*63*
Sdh-1	*28*	*14*	*5*			*43*
Ass-1				*63*	*43*	

Construct a genetic map that describes the locations of all six genes.

C11. If the likelihood of a single crossover in a particular chromosomal region is 10%, what is the theoretical likelihood of a double or triple crossover in that same region? How would positive interference affect these theoretical values?

C12. In most dihybrid crosses involving linked genes, we cannot tell if a double crossover between the two genes has occurred because the offspring inherits the nonrecombinant pattern of alleles. How does the inability to detect double crossovers affect the calculation of map distance? Is map distance underestimated or overestimated because of our inability to detect double crossovers? Explain your answer.

C13. Researchers have discovered that some regions of chromosomes are much more likely than others to cross over. We might call such a region a "hot spot" for crossing over. Let's suppose that two genes, gene *A* and gene *B,* are 5,000,000 bp apart on the same chromosome. Genes *A* and *B* are in a hot spot for crossing over. Two other genes, let's call them gene *C* and gene *D,* are also 5,000,000 bp apart but are not in a hot spot for recombination. If we conducted dihybrid crosses to compute the map distance between genes *A* and *B*, and other dihybrid crosses to compute the map distance between genes *C* and *D,* would the map distances be the same between *A* and *B* compared with to *C* and *D*? Explain.

C14. What is mitotic recombination? A heterozygous individual (*Bb*) with brown eyes has one eye with a small patch of blue. Provide two or more explanations for how the blue patch may have occurred.

C15. Mitotic recombination can occasionally produce a twin spot. Let's suppose an animal is heterozygous for two genes that govern fur color and length: One gene affects pigmentation, with dark pigmentation (*A*) dominant to albino (*a*); the other gene affects hair length, with long hair (*L*) dominant to short hair (*l*). The two genes are linked on the same chromosome. Let's assume an animal is *AaLl*; *A* is linked to *l*, and *a* is linked to *L*. Draw the chromosomes labeled with these alleles, and explain how mitotic recombination could produce a twin spot with one spot having albino pigmentation and long fur, the other having dark pigmentation and short fur.

Application and Experimental Questions (Includes Most Mapping Questions)

E1. Figure 7.1 shows the first experimental results that indicated linkage between two different genes. Conduct a chi square analysis to see if the genes are really linked and the data could not be explained by independent assortment.

E2. In the experiment of Figure 7.6, the researchers followed the inheritance pattern of chromosomes that were abnormal at both ends to correlate genetic recombination with the physical exchange of chromosome pieces. Is it necessary to analyze a chromosome that is abnormal at both ends, or could the researchers have used a parental strain with two abnormal versions of chromosome 9, one with a knob at one end and its homolog with a translocation at the other end?

E3. How would you determine that genes in mammals are located on the Y chromosome linkage group? Is it possible to conduct crosses (let's say in mice) to map the distances between genes along the Y chromosome? Explain.

E4. Explain the rationale behind a testcross. Is it necessary for one of the parents to be homozygous recessive for the genes of interest? In the heterozygous parent of a testcross, must all of the dominant alleles be linked on the same chromosome and all of the recessive alleles be linked on the homolog?

E5. In your own words, explain why a testcross cannot produce more than 50% recombinant offspring. When a testcross does produce 50% recombinant offspring, what do these results mean?

E6. Explain why the percentage of recombinant offspring in a testcross is a more accurate measure of map distance when two genes are close together. When two genes are far apart, is the percentage of recombinant offspring an underestimate or overestimate of the actual map distance?

E7. If two genes are more than 50 mu apart, how would you ever be able to show experimentally that they are located on the same chromosome?

E8. In Morgan's trihybrid testcross of Figure 7.3, he realized that crossing over was more frequent between the eye color and wing length genes than between the body color and eye color genes. Explain how he determined this.

E9. In the experiment of Figure 7.9, list the gene pairs from the particular dihybrid crosses that Sturtevant used to construct his genetic map.

E10. In the tomato, red fruit (*R*) is dominant over yellow fruit (*r*), and yellow flowers (*Wf*) are dominant over white flowers (*wf*). A cross was made between true-breeding plants with red fruit and yellow flowers, and plants with yellow fruit and white flowers. The F_1 generation plants were then crossed to plants with yellow fruit and white flowers. The following results were obtained:

333 red fruit, yellow flowers
64 red fruit, white flowers
58 yellow fruit, yellow flowers
350 yellow fruit, white flowers

Calculate the map distance between the two genes.

E11. Two genes are located on the same chromosome and are known to be 12 mu apart. An *AABB* individual was crossed to an *aabb* individual to produce *AaBb* offspring. The *AaBb* offspring were then crossed to *aabb* individuals.

A. If this cross produces 1000 offspring, what are the predicted numbers of offspring with each of the four genotypes: *AaBb*, *Aabb*, *aaBb*, and *aabb*?

B. What would be the predicted numbers of offspring with these four genotypes if the parental generation had been *AAbb* and *aaBB* instead of *AABB* and *aabb*?

E12. Two genes, designated *A* and *B*, are located 10 mu from each other. A third gene, designated *C*, is located 15 mu from *B* and 5 mu from *A*. The parental generation consisting of *AA bb CC* and *aa BB cc* individuals were crossed to each other. The F_1 heterozygotes were then testcrossed to *aa bb cc* individuals. If we assume no double crossovers occur in this region, what percentage of offspring would you expect with the following genotypes?

A. *Aa Bb Cc*
B. *aa Bb Cc*
C. *Aa bb cc*

E13. Two genes in tomatoes are 61 mu apart; normal fruit (*F*) is dominant to fasciated fruit (*f*), and normal numbers of leaves (*Lf*) is dominant to leafy (*lf*). A true-breeding plant with normal leaves and fruit was crossed to a leafy plant with fasciated fruit. The F_1 offspring were then crossed to leafy plants with fasciated fruit. If this cross produced 600 offspring, what are the expected numbers of plants in each of the four possible categories: normal leaves, normal fruit; normal leaves, fasciated fruit; leafy, normal fruit; and leafy, fasciated fruit?

E14. In the tomato, three genes are linked on the same chromosome. Tall is dominant to dwarf, skin that is smooth is dominant to skin that is peachy, and fruit with a normal tomato shape is dominant to oblate shape. A plant that is true-breeding for the dominant traits was crossed to a dwarf plant with peachy skin and oblate fruit. The F_1 plants were then testcrossed to dwarf plants with peachy skin and oblate fruit. The following results were obtained:

151 tall, smooth, normal
33 tall, smooth, oblate
11 tall, peach, oblate
2 tall, peach, normal
155 dwarf, peach, oblate
29 dwarf, peach, normal
12 dwarf, smooth, normal
0 dwarf, smooth, oblate

Construct a genetic map that describes the order of these three genes and the distances between them.

E15. A trait in garden peas involves the curling of leaves. A dihybrid cross was made involving a plant with yellow pods and curling leaves to a wild-type plant with green pods and normal leaves. All F_1 offspring had green pods and normal leaves. The F_1 plants were then crossed to plants with yellow pods and curling leaves. The following results were obtained:

117 green pods, normal leaves
115 yellow pods, curling leaves
78 green pods, curling leaves
80 yellow pods, normal leaves

A. Conduct a chi square analysis to determine if these two genes are linked.

B. If they are linked, calculate the map distance between the two genes. How accurate do you think this distance is?

E16. In mice, the gene that encodes the enzyme inosine triphosphatase is 12 mu from the gene that encodes the enzyme ornithine decarboxylase. Suppose you have identified a strain of mice homozygous for a defective inosine triphosphatase gene that does not produce any of this enzyme and is also homozygous for a defective ornithine decarboxylase gene. In other words, this strain of mice cannot make either enzyme. You crossed this homozygous recessive strain to a normal strain of mice to produce heterozygotes. The heterozygotes were then backcrossed to the strain that cannot produce either enzyme. What is the probability of obtaining a mouse that cannot make either enzyme?

E17. In the garden pea, several different genes affect pod characteristics. A gene affecting pod color (green is dominant to yellow) is approximately 7 mu away from a gene affecting pod width (wide is dominant to narrow). Both genes are located on chromosome 5. A third gene, located on chromosome 4, affects pod length (long is dominant to short). A true-breeding wild-type plant (green, wide, long pods) was crossed to a plant with yellow, narrow, short pods. The F_1 offspring were then testcrossed to plants with yellow, narrow, short pods. If the testcross produced 800 offspring, what are the expected numbers of the eight possible phenotypic combinations?

E18. A sex-influenced trait is dominant in males and causes bushy tails. The same trait is recessive in females. Fur color is not sex-influenced. Yellow fur is dominant to white fur. A true-breeding female with a bushy tail and yellow fur was crossed to a white male without a bushy tail (i.e., a normal tail). The F_1 females were then crossed to white males without bushy tails. The following results were obtained:

Males	**Females**
28 normal tails, yellow	102 normal tails, yellow
72 normal tails, white	96 normal tails, white
68 bushy tails, yellow	0 bushy tails, yellow
29 bushy tails, white	0 bushy tails, white

A. Conduct a chi square analysis to determine if these two genes are linked.

B. If the genes are linked, calculate the map distance between them. Explain which data you used in your calculation.

E19. Three recessive traits in garden pea plants are as follows: yellow pods are recessive to green pods, bluish green seedlings are recessive to green seedlings, creeper (a plant that cannot stand up) is recessive to normal. A true-breeding normal plant with green pods and green seedlings was crossed to a creeper with yellow pods and bluish green seedlings. The F_1 plants were then crossed to creepers with yellow pods and bluish green seedlings. The following results were obtained:

2059 green pods, green seedlings, normal
151 green pods, green seedlings, creeper
281 green pods, bluish green seedlings, normal
15 green pods, bluish green seedlings, creeper
2041 yellow pods, bluish green seedlings, creeper
157 yellow pods, bluish green seedlings, normal
282 yellow pods, green seedlings, creeper
11 yellow pods, green seedlings, normal

Construct a genetic map that describes the map distance between these three genes.

E20. In mice, a trait called snubnose is recessive to a wild-type nose, a trait called pintail is dominant to a normal tail, and a trait called jerker (a defect in motor skills) is recessive to a normal gait. Jerker mice with a snubnose and pintail were crossed to normal mice, and then the F_1 mice were crossed to jerker mice that have a snubnose and normal tail. The outcome of this cross was as follows:

560 jerker, snubnose, pintail
548 normal gait, normal nose, normal tail
102 jerker, snubnose, normal tail
104 normal gait, normal nose, pintail

77 jerker, normal nose, normal tail
71 normal gait, snubnose, pintail
11 jerker, normal nose, pintail
9 normal gait, snubnose, normal tail

Construct a genetic map that describes the order and distance between these genes.

E21. In *Drosophila,* an allele causing vestigial wings is 12.5 mu away from another allele that causes purple eyes. A third gene that affects body color has an allele that causes black body color. This third gene is 18.5 mu away from the vestigial wings allele and 6 mu away from the allele causing purple eyes. The alleles causing vestigial wings, purple eyes, and black body are all recessive. The dominant (wild-type) traits are long wings, red eyes, and gray body. A researcher crossed wild-type flies to flies with vestigial wings, purple eyes, and black bodies. All F_1 flies were wild type. F_1 female flies were then crossed to male flies with vestigial wings, purple eyes, and black bodies. If 1000 offspring were observed, what are the expected numbers of the following types of flies?

Long wings, red eyes, gray body
Long wings, purple eyes, gray body
Long wings, red eyes, black body
Long wings, purple eyes, black body
Short wings, red eyes, gray body
Short wings, purple eyes, gray body
Short wings, red eyes, black body
Short wings, purple eyes, black body

Which kinds of flies can be produced only by a double crossover event?

E22. Three autosomal genes are linked along the same chromosome. The distance between gene *A* and *B* is 7 mu, the distance between *B* and *C* is 11 mu, and the distance between *A* and *C* is 4 mu. An individual who is *AA bb CC* was crossed to an individual who is *aa BB cc* to produce heterozygous F_1 offspring. The F_1 offspring were then crossed to homozygous *aa bb cc* individuals to produce F_2 offspring.

A. Draw the arrangement of alleles on the chromosomes in the parents and in the F_1 offspring.

B. Where would a crossover have to occur to produce an F_2 offspring that was heterozygous for all three genes?

C. If we assume that no double crossovers occur in this region, what percentage of F_2 offspring is likely to be homozygous for all three genes?

E23. Let's suppose that two different X-linked genes exist in mice, designated with the letters *N* and *L*. Gene *N* exists in a dominant, normal allele and in a recessive allele, *n*, that is lethal. Similarly, gene *L* exists in a dominant, normal allele and in a recessive allele, *l*, that is lethal. Heterozygous females are normal, but males that carry either recessive allele are born dead. Explain whether or not it would be possible to map the distance between these two genes by making crosses and analyzing the number of living and dead offspring. You may assume that you have strains of mice in which females are heterozygous for one or both genes.

Questions for Student Discussion/Collaboration

1. In mice, a dominant allele that causes a short tail is located on chromosome 2. On chromosome 3, a recessive allele causing droopy ears is 6 mu away from another recessive allele that causes a flaky tail. A recessive allele that causes a jerker (uncoordinated) phenotype is located on chromosome 4. A jerker mouse with droopy ears and a short, flaky tail was crossed to a normal mouse. All F_1 generation mice were phenotypically normal, except they had short tails. These F_1 mice were then testcrossed to jerker mice with droopy ears and long, flaky tails. If this cross produced 400 offspring, what would be the proportions of the 16 possible phenotypic categories?
2. In Chapter 4, we discussed the idea that the X and Y chromosomes have a few genes in common. These genes are inherited in a pseudoautosomal pattern. With this phenomenon in mind, discuss whether or not the X and Y chromosomes are really distinct linkage groups.
3. Mendel studied seven traits in pea plants, and the garden pea happens to have seven different chromosomes. It has been pointed out that Mendel was very lucky not to have conducted crosses involving two traits governed by genes that are closely linked on the same chromosome because the results would have confounded his theory of independent assortment. It has even been suggested that Mendel may not have published data involving traits that were linked! An article by Stig Blixt ("Why Didn't Gregor Mendel Find Linkage?" *Nature 256*:206, 1975) considers this issue. Look up this article and discuss why Mendel did not find linkage.

Answers to Comprehension Questions

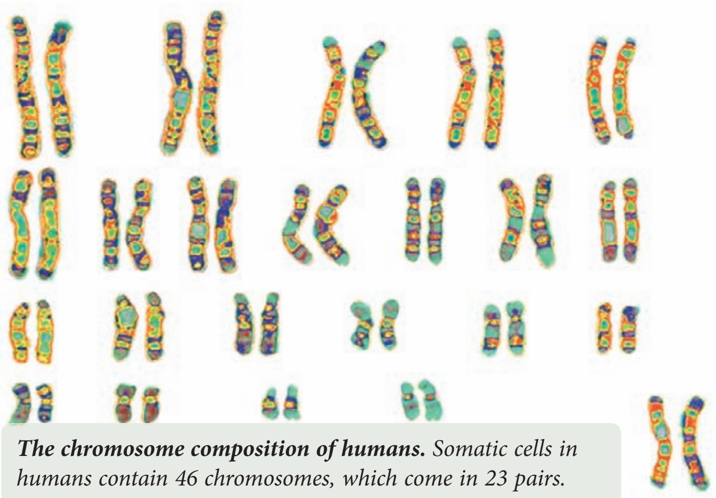
7.1: b, d
7.2: a, a, b
7.3: d, d, b
7.4: b

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

8



The chromosome composition of humans. *Somatic cells in humans contain 46 chromosomes, which come in 23 pairs.*

CHAPTER OUTLINE

8.1 Microscopic Examination of Eukaryotic Chromosomes

8.2 Changes in Chromosome Structure: An Overview

8.3 Deletions and Duplications

8.4 Inversions and Translocations

8.5 Changes in Chromosome Number: An Overview

8.6 Variation in the Number of Chromosomes Within a Set: Aneuploidy

8.7 Variation in The Number of Sets of Chromosomes

8.8 Mechanisms That Produce Variation in Chromosome Number

VARIATION IN CHROMOSOME STRUCTURE AND NUMBER

The term **genetic variation** refers to genetic differences among members of the same species or between different species. Throughout Chapters 2 to 7, we have focused primarily on variation in specific genes, which is called **allelic variation.** In Chapter 8, our emphasis will shift to larger types of genetic changes that affect the structure or number of eukaryotic chromosomes. These larger alterations may affect the expression of many genes, thereby influencing phenotypes. Variation in chromosome structure and number are of great importance in the field of genetics because they are critical in the evolution of new species and have widespread medical relevance. These topics are discussed in Chapters 27 and 23, respectively. In addition, agricultural geneticists have discovered that such variation can lead to the development of new strains of crops, which may be quite profitable.

In the first half of Chapter 8, we will examine how the structure of a eukaryotic chromosome can be modified, either by altering the total amount of genetic material or by rearranging the order of genes along a chromosome. Such changes may often be detected microscopically. The rest of the chapter is concerned with changes in the total number of chromosomes. We will explore how variation in chromosome number occurs and consider examples in which it has significant phenotypic consequences. We will also examine how changes in chromosome number can be induced through experimental treatments and how these approaches have applications in agriculture.

8.1 MICROSCOPIC EXAMINATION OF EUKARYOTIC CHROMOSOMES

Chromosomes in the nuclei of eukaryotic cells contain long, linear DNA molecules that carry hundreds or even thousands of genes. In this chapter, we will consider how the structure and number of chromosomes can be changed. A **cytogeneticist** is a scientist who studies the structure and function of the chromosomes with the aid of a microscope. Because chromosomes are more compact and readily visible during cell division, a cytogeneticist usually analyzes chromosomes in actively dividing cells. To determine what the normal chromosomes of a species look like, the chromosomes from several members of a given species are examined. In most cases, two phenotypically normal individuals of the same species have the same number and types of chromosomes.

Figure 8.1a shows micrographs of chromosomes from three species: a human, a fruit fly, and a corn plant. As seen here, a human has 46 chromosomes (23 pairs), a fruit fly has 8 chromosomes (4 pairs), and corn has 20 chromosomes (10 pairs). Except for the sex chromosomes, which differ between males and females, most members of the same species have very similar chromosomes. For example, the majority of people have 46 chromosomes in their somatic cells. By comparison, the chromosomal compositions of distantly related species, such as humans and fruit flies, may be very different. A total of 46 chromosomes is normal for humans, whereas 8 chromosomes is the norm for fruit flies.

Cytogeneticists have various ways to classify and identify chromosomes. The three most commonly used features are location of the centromere, size, and banding patterns that are revealed when the chromosomes are treated with stains. As shown in **Figure 8.1b**, chromosomes are classified as **metacentric** (in which the centromere is near the middle), **submetacentric** (in which the centromere is slightly off center), **acrocentric** (in which the centromere is significantly off center but not at the end), and **telocentric** (in which the centromere is at one end). Because the centromere is never exactly in the center of a chromosome, each chromosome has a short arm and a long arm. For human chromosomes, the short arm is designated with the letter *p* (for the French, petite), and the long arm is designated with the letter *q*. In the case of telocentric chromosomes, the short arm may be nearly nonexistent.

Figure 8.1c shows a human karyotype. The procedure for making a karyotype is described in Chapter 2 (see Figure 2.2a). A **karyotype** is a micrograph in which all of the chromosomes within a single cell have been arranged in a standard fashion. When preparing a karyotype, the chromosomes are aligned with the short arms on top and the long arms on the bottom. By convention, the chromosomes are numbered roughly according to their size, with the largest chromosomes having the smallest numbers. For example, human chromosomes 1, 2, and 3 are relatively large, whereas 21 and 22 are the two smallest. An exception to the numbering system involves the sex chromosomes, which are designated with letters (for humans, X and Y).

Because different chromosomes often have similar sizes and centromeric locations (e.g., compare human chromosomes 8, 9, and 10), geneticists must use additional methods to accurately identify each type of chromosome within a karyotype. For detailed identification, chromosomes are treated with stains to produce characteristic banding patterns. Several different staining procedures are used by cytogeneticists to identify specific chromosomes. An example is **G bands,** which are shown in Figure 8.1c. In this procedure, chromosomes are treated with mild heat or with proteolytic enzymes that partially digest chromosomal proteins. When exposed to the dye called Giemsa, named after its inventor Gustav Giemsa, some chromosomal regions bind the dye heavily and produce a dark band. In other regions, the stain hardly binds at all and a light band results. Though the mechanism of staining is not completely understood, the dark bands are thought to represent regions that are more tightly compacted. As shown in Figure 8.1c and d, the alternating pattern of G bands is a unique feature for each chromosome.

In the case of human chromosomes, approximately 300 G bands can usually be distinguished during metaphase. A larger number of G bands (in the range of 800) can be observed in prometaphase chromosomes because they are more extended than metaphase chromosomes. **Figure 8.1d** shows the conventional numbering system that is used to designate G bands along a set of human chromosomes. The left chromatid in each pair of sister chromatids shows the expected banding pattern during metaphase, and the right chromatid shows the banding pattern as it would appear during prometaphase.

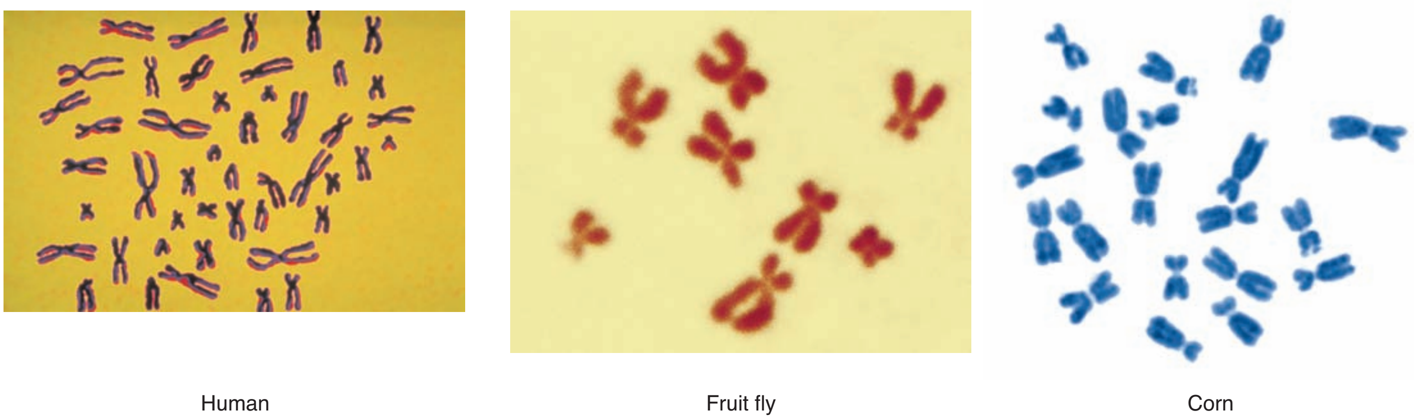
Why is the banding pattern of eukaryotic chromosomes useful? First, when stained, individual chromosomes can be distinguished from each other, even if they have similar sizes and centromeric locations. For example, compare the differences in banding patterns between human chromosomes 8 and 9 (Figure 8.1d). These differences permit us to distinguish these two chromosomes even though their sizes and centromeric locations are very similar. Banding patterns are also used to detect changes in chromosome structure. As discussed next, chromosomal rearrangements or changes in the total amount of genetic material are more easily detected in banded chromosomes. Also, chromosome banding can be used to assess evolutionary relationships between species. Research studies have shown that the similarity of chromosome banding patterns is a good measure of genetic relatedness.

8.1 REVIEWING THE KEY CONCEPTS

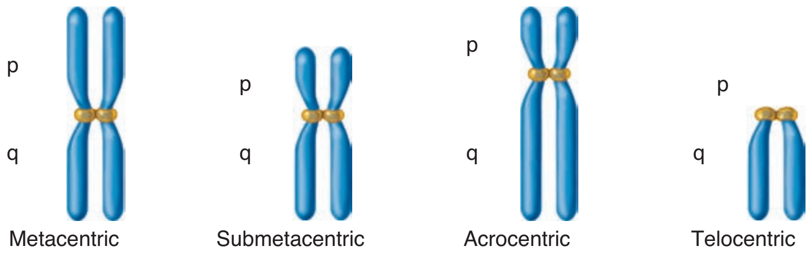
- Among different species, natural variation exists with regard to chromosome structure and number. Three features of chromosomes that aid in their identification are centromere location, size, and banding patterns (see Figure 8.1).

8.1 COMPREHENSION QUESTIONS

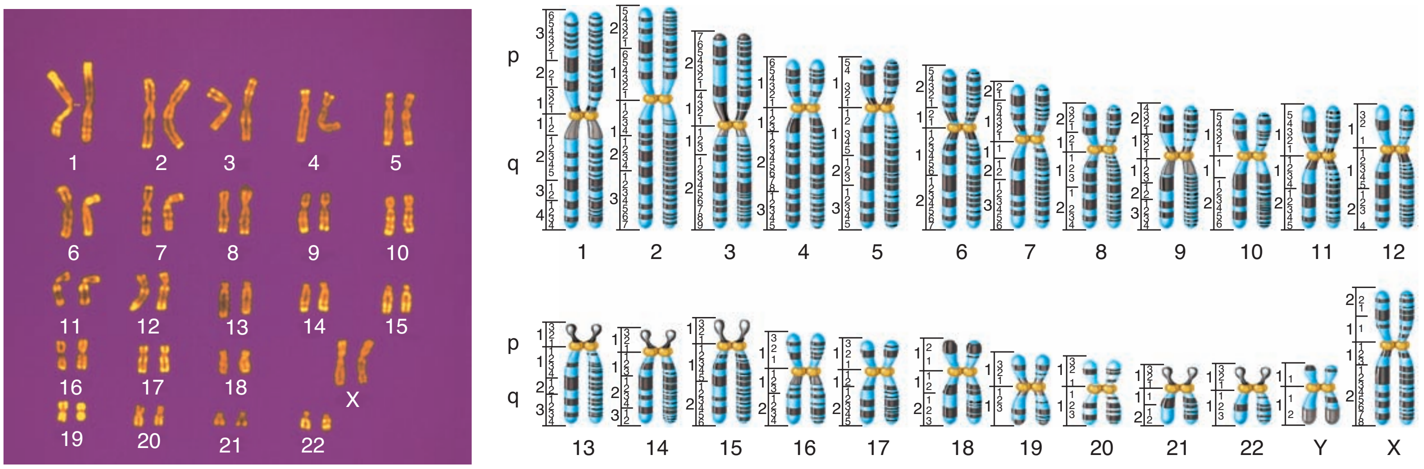
1. A chromosome that is metacentric has its centromere
   a. at the very tip.
   b. near one end, but not at the very tip.
   c. near the middle.
   d. at two distinct locations.
2. Staining eukaryotic chromosomes is useful because it makes it possible to
   a. distinguish chromosomes that are similar in size and centromeric locations.
   b. identify changes in chromosome structure.
   c. explore evolutionary relationships among different species.
   d. do all of the above.


Human
Fruit fly
Corn

(a) Micrographs of metaphase chromosomes


p
q
Metacentric
p
q
Submetacentric
p
q
Acrocentric
p
q
Telocentric

(b) A comparison of centromeric locations


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
19
20
21
22
p
q
1
2
3
4
5
6
7
8
9
10
11
12
p
q
13
14
15
16
17
18
19
20
21
22
Y
X

(c) G banding of human chromosomes

(d) Conventional numbering system of G bands in human chromosomes

FIGURE 8.1 **Features of normal chromosomes.** **(a)** Micrographs of chromosomes from a human, a fruit fly, and corn. **(b)** A comparison of centromeric locations. Centromeres can be metacentric, submetacentric, acrocentric (near one end), or telocentric (at the end). **(c)** Human chromosomes that have been stained with Giemsa. The micrograph has been colorized to more easily discern the banding patterns. **(d)** The conventional numbering of bands in Giemsa-stained human chromosomes. The numbering is divided into broad regions, which then are subdivided into smaller regions. The numbers increase as the region gets farther away from the centromere. For example, if you take a look at the left chromatid of chromosome 1, the uppermost dark band is at a location designated p35. The banding patterns of chromatids change as the chromatids condense. The left chromatid of each pair of sister chromatids shows the banding pattern of a chromatid in metaphase, and the right side shows the banding pattern as it would appear in prometaphase. *Note:* In prometaphase, the chromatids are more extended than in metaphase.

Concept check: Why is it useful to stain chromosomes?

8.2 CHANGES IN CHROMOSOME STRUCTURE: AN OVERVIEW

Now that we understand that chromosomes typically come in a variety of shapes and sizes, let's consider how the structures of normal chromosomes can be modified by mutation. In some cases, the total amount of genetic material within a single chromosome can be increased or decreased significantly. Alternatively, the genetic material in one or more chromosomes may be rearranged without affecting the total amount of material. As shown in **Figure 8.2**, these mutations are categorized as deletions, duplications, inversions, and translocations.

Deletions and duplications are changes in the total amount of genetic material within a single chromosome. In Figure 8.2, human chromosomes are labeled according to their normal G-banding patterns. When a **deletion** occurs, a segment of chromosomal material is missing. In other words, the affected chromosome is deficient in a significant amount of genetic material. The term **deficiency** is also used to describe a missing region of a chromosome. In contrast, a **duplication** occurs when a section of a chromosome is repeated compared with the normal parent chromosome.

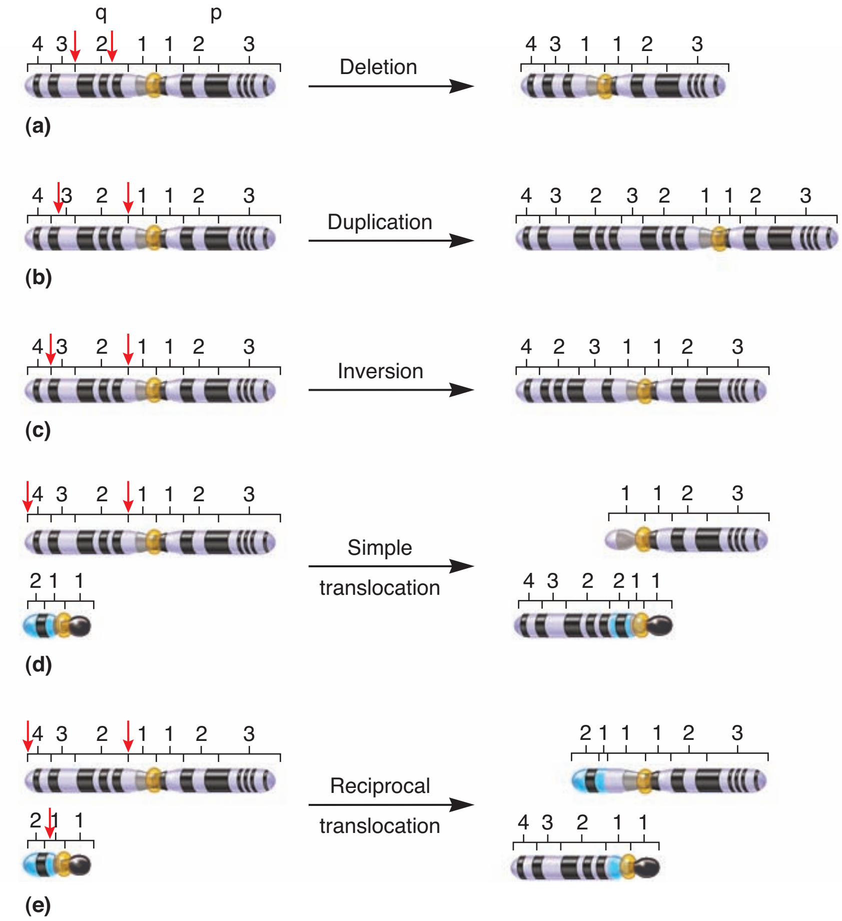
Inversions and translocations are chromosomal rearrangements. An **inversion** involves a change in the direction of the genetic material along a single chromosome. For example, in Figure 8.2c, a segment of one chromosome has been inverted, so the order of four G bands is opposite to that of the parent chromosome. A **translocation** occurs when one segment of a chromosome becomes attached to a different chromosome or to a different part of the same chromosome. A **simple translocation** occurs when a single piece of chromosome is attached to another chromosome. In a **reciprocal translocation,** two different types of chromosomes exchange pieces, thereby producing two abnormal chromosomes carrying translocations.

8.2 REVIEWING THE KEY CONCEPTS

- Variations in chromosome structure include deletions, duplications, inversions, and translocations (see Figure 8.2).

8.2 COMPREHENSION QUESTIONS

1. A change in chromosome structure that does *not* involve a change in the total amount of genetic material is
   a. a deletion.
   b. a duplication.
   c. an inversion.
   d. none of the above.



ONLINE ANIMATION

FIGURE 8.2 Types of changes in chromosome structure. The large chromosome shown throughout is human chromosome 1. The smaller chromosome seen in (d) and (e) is human chromosome 21. The red arrows refer to the ends of the affected portion. **(a)** A deletion occurs that removes a large portion of the q2 region. **(b)** A duplication occurs that doubles the q2-q3 region. **(c)** An inversion occurs that inverts the q2-q3 region. **(d)** The q2-q4 region of chromosome 1 is translocated to chromosome 21. A region of a chromosome cannot be attached directly to the tip of another chromosome because telomeres at the tips of chromosomes prevent such an event. In this example, a small piece at the end of chromosome 21 must be removed for the q2-q4 region of chromosome 1 to be attached to chromosome 21. **(e)** The q2-q4 region of chromosome 1 is exchanged with most of the q1-q2 region of chromosome 21.

Concept check: Which of these changes alter the total amount of genetic material?

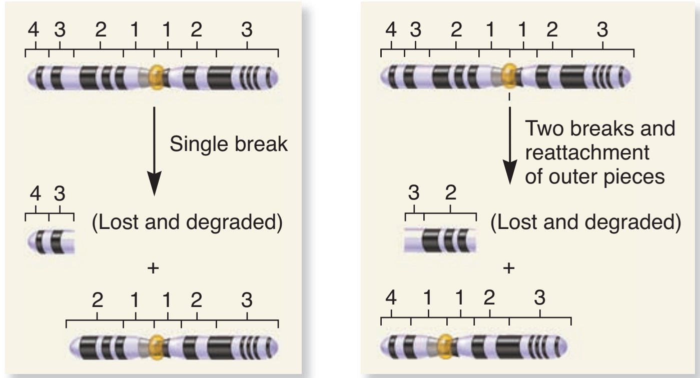
8.3 DELETIONS AND DUPLICATIONS

As we have seen, deletions and duplications alter the total amount of genetic material within a chromosome. In this section, we will examine how these changes occur, how the changes are detected experimentally, and how they affect the phenotypes of the individuals who inherit them.

The Loss of Genetic Material in a Deletion Tends to Be Detrimental to an Organism

A chromosomal deletion occurs when a chromosome breaks in one or more places and a fragment of the chromosome is lost. In **Figure 8.3a**, a normal chromosome has broken into two separate pieces. The piece without the centromere is lost and degraded. This event produces a chromosome with a **terminal deletion.** In **Figure 8.3b**, a chromosome has broken in two places to produce three chromosomal fragments. The central fragment is lost, and the two outer pieces reattach to each other. This process has created a chromosome with an **interstitial deletion.** Deletions can also be created when recombination takes place at incorrect locations between two homologous chromosomes. The products of this type of aberrant recombination event are one chromosome with a deletion and another chromosome with a duplication. This process is examined later in this chapter.

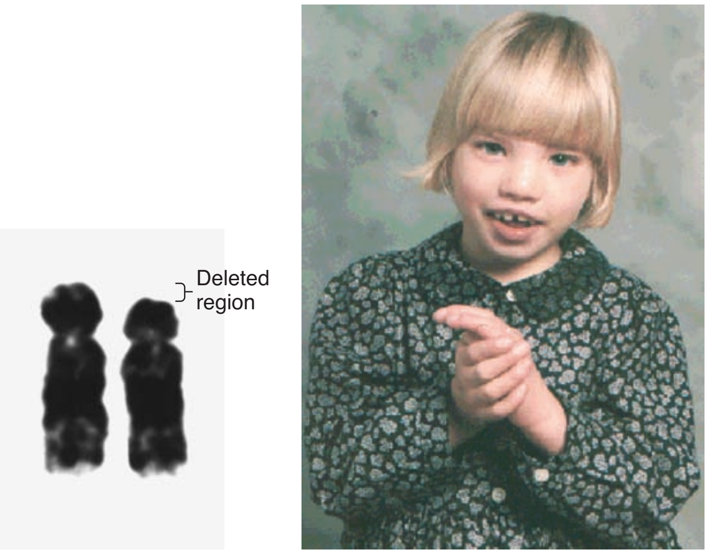
The phenotypic consequences of a chromosomal deletion depend on the size of the deletion and whether it includes genes or portions of genes that are vital to the development of the organism. When deletions have a phenotypic effect, they are usually detrimental. Larger deletions tend to be more harmful because more genes are missing. Many examples are known in which deletions have significant phenotypic influences. For example, a human genetic disease known as cri-du-chat, or Lejeune, syndrome is caused by a deletion in a segment of the short arm of human chromosome 5 (**Figure 8.4a**). Individuals who carry a single copy of this abnormal chromosome along with a normal chromosome 5 display an array of abnormalities, including mental deficiencies, unique facial anomalies, and an unusual catlike cry in infancy, which is the meaning of the French name for the syndrome (**Figure 8.4b**). Other human genetic diseases, such as Angelman syndrome and Prader-Willi syndrome, described in Chapter 6, are due to a deletion in chromosome 15.



(a) Terminal deletion **(b) Interstitial deletion**

FIGURE 8.3 **Production of terminal and interstitial deletions.** This illustration shows the production of deletions in human chromosome 1.

Concept check: What is the reason why a chromosomal fragment is lost and then subsequently degraded?



(a) Chromosome 5 **(b) A child with cri-du-chat syndrome**

FIGURE 8.4 **Cri-du-chat syndrome. (a)** Chromosome 5 from the karyotype of an individual with this disorder. A section of the short arm of chromosome 5 is missing. **(b)** An affected individual.

Genes → Traits Compared with an individual who has two copies of each gene on chromosome 5, an individual with cri-du-chat syndrome has only one copy of the genes that are located within the missing segment. This genetic imbalance (one versus two copies of many genes on chromosome 5) causes the phenotypic characteristics of this disorder, which include a catlike cry in infancy, short stature, characteristic facial anomalies (e.g., a triangular face, almond-shaped eyes, broad nasal bridge, and low-set ears), and microencephaly (a smaller-than-normal brain).

Duplications Tend to Be Less Harmful Than Deletions

Duplications result in extra genetic material. They are usually caused by abnormal events during recombination. Under normal circumstances, crossing over occurs at properly aligned sites between homologous chromosomes. On rare occasions, a crossover may occur at misaligned sites on the homologs (**Figure 8.5**). What causes the misalignment? In some cases, a chromosome may carry two or more homologous segments of DNA that have identical or similar sequences. These are called **repetitive sequences** because they occur multiple times. An example of repetitive sequences is transposable elements, which are described in Chapter 21. In Figure 8.5, the repetitive sequence on the right (in the upper chromatid) has lined up with the repetitive sequence on the left (in the lower chromatid). A crossover then occurs. This is called **nonallelic homologous recombination** because it has occurred at homologous sites (i.e., repetitive sequences), but the alleles of

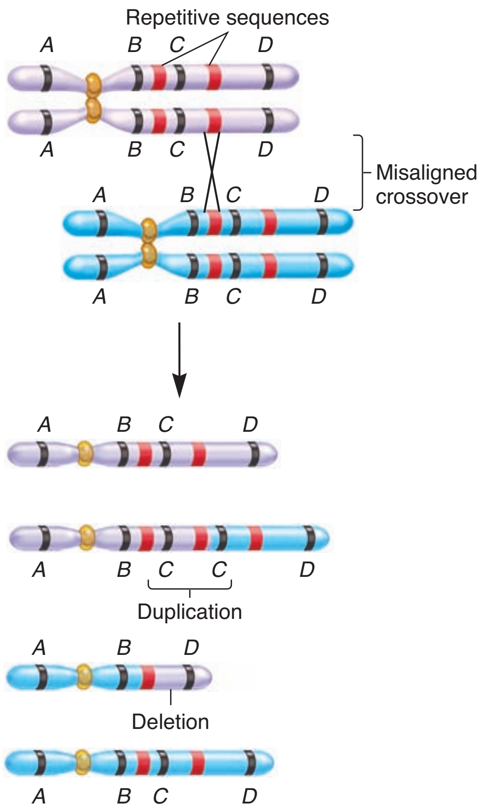


FIGURE 8.5 **Nonallelic homologous recombination, leading to a duplication and a deletion.** Repetitive sequences (shown in red) have promoted the misalignment of homologous chromosomes. A crossover has occurred at sites between genes *C* and *D* in one chromatid and between genes *B* and *C* in another chromatid. After crossing over is completed, one chromatid contains a duplication, and the other contains a deletion.

Concept check: *What is the underlying cause of nonallelic homologous recombination?*

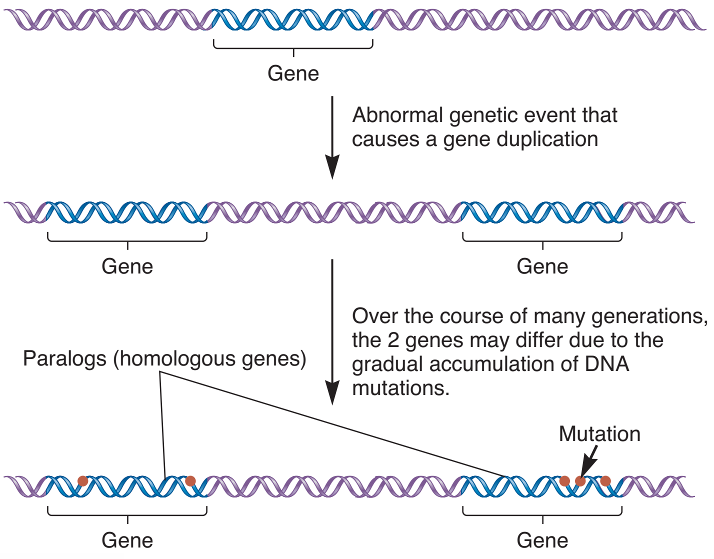
neighboring genes are not properly aligned. The result is that one chromatid has an internal duplication and another chromatid has a deletion. In Figure 8.5, the chromosome with the extra genetic material carries a **gene duplication,** because the number of copies of gene *C* has been increased from one to two. In most cases, gene duplications happen as rare, sporadic events during the evolution of species. Later in this section, we will consider how multiple copies of genes can evolve into a family of genes with specialized functions.

Like deletions, the phenotypic consequences of duplications tend to be correlated with size. Duplications are more likely to have phenotypic effects if they involve a large piece of the chromosome. In general, small duplications are less likely to have harmful effects than are deletions of comparable size. This observation suggests that having only one copy of a gene is more harmful than having three copies. In humans, relatively few well-defined syndromes are caused by small chromosomal duplications. An example is Charcot-Marie-Tooth disease (type 1A), a peripheral neuropathy characterized by numbness in the hands and feet that is caused by a small duplication on the short arm of chromosome 17.

Duplications Provide Additional Material for Gene Evolution, Sometimes Leading to the Formation of Gene Families

In contrast to the gene duplication that causes Charcot-Marie-Tooth disease, the majority of small chromosomal duplications have no phenotypic effect. Nevertheless, they are vitally important because they provide raw material for the addition of more genes into a species' chromosomes. Over the course of many generations, this can lead to the formation of a **gene family** consisting of two or more genes that are similar to each other. As shown in **Figure 8.6**, the members of a gene family are derived from the same ancestral gene. Over time, two copies of an ancestral gene can accumulate different mutations. Therefore, after many generations, the two genes are similar but not identical. During evolution, this type of event can occur several times, creating a family of many similar genes.

When two or more genes are derived from a single ancestral gene, the genes are said to be **homologous.** Homologous genes within a single species are called **paralogs** and constitute a gene family. A well-studied example of a gene family is shown in **Figure 8.7**, which illustrates the evolution of the globin gene family found in humans. The globin genes encode polypeptides that are subunits of proteins that function in oxygen binding. For example, hemoglobin is a protein found in red blood cells; its function is to carry oxygen throughout the body. The globin gene family is composed of 14 paralogs that were originally derived from a single ancestral globin gene. According to an evolutionary analysis, the ancestral globin gene first duplicated about 500 million years ago and became separate genes encoding myoglobin and the hemoglobin group of genes. The primordial hemoglobin gene duplicated into an α-chain gene and a β-chain gene, which subsequently duplicated to produce several genes located on chromosomes 16 and 11, respectively. Currently, 14 globin genes are found on three different human chromosomes.


Gene
Abnormal genetic event that causes a gene duplication
Gene
Gene
Over the course of many generations, the 2 genes may differ due to the gradual accumulation of DNA mutations.
Paralogs (homologous genes)
Mutation
Gene
Gene

ONLINE ANIMATION

FIGURE 8.6 **Gene duplication and the evolution of paralogs.** An abnormal crossover event like the one described in Figure 8.5 leads to a gene duplication. Over time, each gene accumulates different mutations.

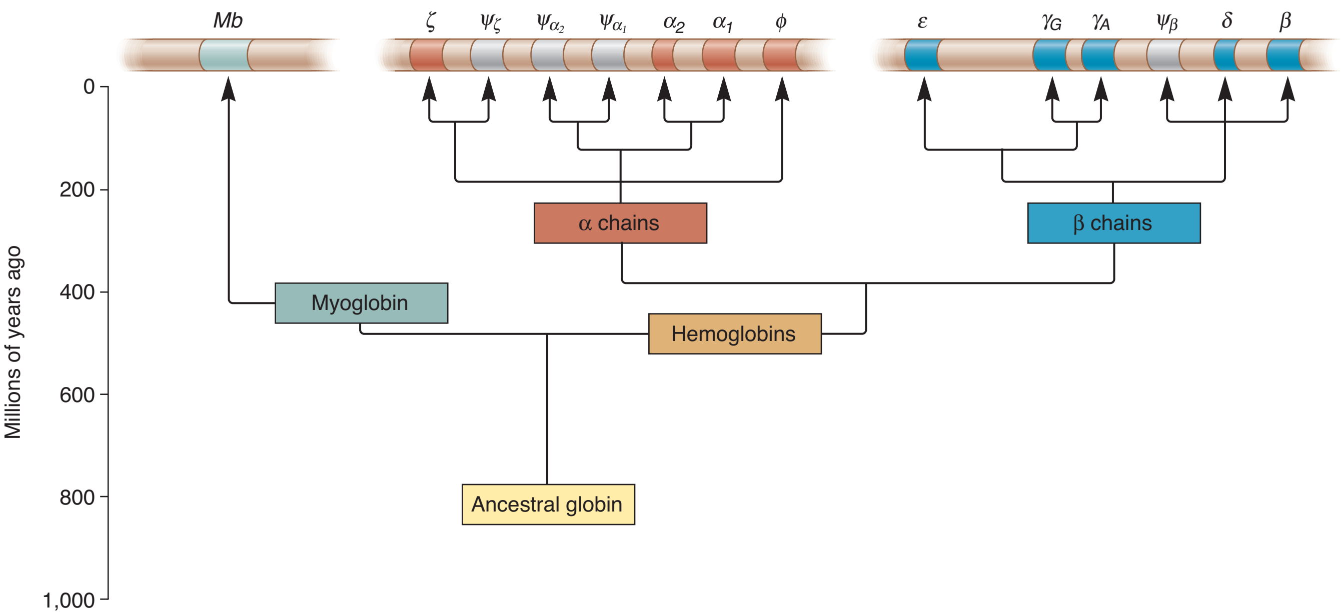


FIGURE 8.7 **The evolution of the globin gene family in humans.** The globin gene family evolved from a single ancestral globin gene. The first gene duplication produced two genes that accumulated mutations and became the genes encoding myoglobin (on chromosome 22) and the group of hemoglobins. The primordial hemoglobin gene then duplicated to produce several α-chain and β-chain genes, which are found on chromosomes 16 and 11, respectively. The four genes shown in gray are nonfunctional pseudogenes.

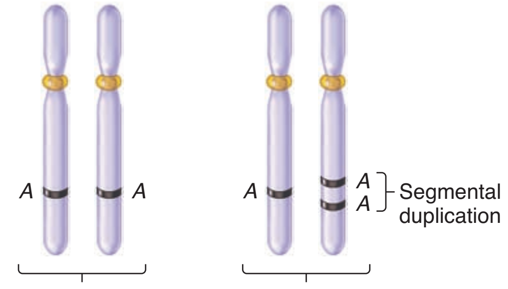
Why is it advantageous to have a family of globin genes? Although all globin polypeptides are subunits of proteins that play a role in oxygen binding, the accumulation of different mutations in the various family members has produced globins that are more specialized in their function. For example, myoglobin is better at binding and storing oxygen in muscle cells, and the hemoglobins are better at binding and transporting oxygen via the red blood cells. Also, different globin genes are expressed during different stages of human development. The ε- and ζ-globin genes are expressed very early in embryonic life, whereas the α-globin and γ-globin genes are expressed during the second and third trimesters of gestation. Following birth, the α-globin gene remains turned on, but the γ-globin genes are turned off and the β-globin gene is turned on. These differences in the expression of the globin genes reflect the differences in the oxygen transport needs of humans during the embryonic, fetal, and postpartum stages of life.

Copy Number Variation Is Relatively Common Among Members of the Same Species

The term **copy number variation (CNV)** refers to a type of structural variation in which a segment of DNA that is 1000 bp or more in length exhibits copy number differences among members of the same species. Genes that were assumed to always occur in two copies in a diploid cell have now been found to be present in one, three, or more than three copies. One possibility is that some members of a species may carry a chromosome that is missing a particular gene or part of a gene. Alternatively, a CNV may involve a duplication. For example, some members of a diploid species may have one copy of gene *A* on both homologs of a chromosome, and thereby have two copies of the gene (**Figure 8.8a**). By comparison, other members of the same species might have one copy of gene *A* on a particular chromosome and two copies on its homolog for a total of three copies (**Figure 8.8b**). The homolog with two copies of gene *A* is said to have undergone a **segmental duplication.**

In the past 10 years, researchers have discovered that copy number variation is relatively common in animal and plant species. Though the analysis of structural variation is a relatively new area of investigation, researchers estimate that 1% to 10% of a genome may show CNV within a typical species of animal or plant.

Most CNV is inherited and has happened in the past, but it may also be caused by new mutations. A variety of mechanisms may bring about copy number variation. One common cause



(a) Some members of a species

(b) Other members of the same species

FIGURE 8.8 **An example of copy number variation.** In part (**a**), an individual has two copies of gene *A*, whereas another individual in part (**b**) has three copies.

is nonallelic homologous recombination, which was described in Figure 8.5. This type of event can produce a chromosome with a duplication or deletion, thereby altering the copy number of genes. Researchers also speculate that the proliferation of transposable elements, which are described in Chapter 21, may increase the copy number of DNA segments. A third mechanism that underlies CNV may involve errors in DNA replication, which is described in Chapter 13.

What are the phenotypic consequences of CNV? In many cases, CNV has no obvious phenotypic consequences. However, recent medical research is revealing that some CNV in humans is associated with specific diseases. For example, particular types of CNV are associated with schizophrenia, autism, and certain forms of learning disabilities. In addition, CNV may affect susceptibility to infectious diseases. An example is the human *CCL3* gene that encodes a chemokine protein, which is involved in immunity. In human populations, the copy number of this gene varies from one to six. In people infected with HIV (human immunodeficiency virus), copy number variation of *CCL3* may affect the progression of AIDS (acquired immune deficiency syndrome). Individuals with a higher copy number of *CCL3* produce more chemokine protein and often show a slower advancement of AIDS. Finally, another reason why researchers are interested in copy number variation is its possible relationship to cancer, which is discussed in Chapter 23.

8.3 REVIEWING THE KEY CONCEPTS

- Chromosome breaks can create terminal and interstitial deletions. Some deletions are associated with human genetic disorders such as cri-du-chat syndrome (see Figures 8.3, 8.4).
- Nonallelic homologous recombination can create gene duplications and deletions. Over time, gene duplications can lead to the formation of gene families, such as the globin gene family (see Figures 8.5–8.7).
- Copy number variation (CNV) is fairly common within a species. In humans, CNV is associated with certain diseases (see Figure 8.8).

8.3 COMPREHENSION QUESTIONS

1. Which of the following statements is correct?
   a. If a deletion and duplication are the same size, the deletion is more likely to be harmful.
   b. If a deletion and duplication are the same size, the duplication is more likely to be harmful.
   c. If a deletion and duplication are the same size, the likelihood of causing harm is about the same.
   d. A deletion is always harmful whereas a duplication is always beneficial.
2. With regard to gene duplications, which of the following statement(s) is/are correct?
   a. Gene duplications may be caused by nonallelic homologous recombination.
   b. Large gene duplications are more likely to be harmful than smaller ones.
   c. Gene duplications are responsible for creating gene families that encode proteins with similar and specialized functions.
   d. All of the above statements are correct.

8.4 INVERSIONS AND TRANSLOCATIONS

As discussed earlier in this chapter, inversions and translocations are types of chromosomal rearrangements. In this section, we will explore how they occur and how they may affect an individual's phenotype or fertility.

Inversions Often Occur Without Phenotypic Consequences

A chromosome with an inversion contains a segment that has been flipped to the opposite direction (**Figure 8.9**). Geneticists classify inversions according to the location of the centromere. If the centromere lies within the inverted region of the chromosome, the inverted region is known as a **pericentric inversion** (Figure 8.9b). Alternatively, if the centromere is found outside the inverted region, the inverted region is called a **paracentric inversion** (Figure 8.9c).

When a chromosome contains an inversion, the total amount of genetic material remains the same as in a normal chromosome. Therefore, the great majority of inversions do not have any phenotypic consequences. In rare cases, however, an inversion can alter the phenotype of an individual. Whether or not this occurs is related to the boundaries of the inverted segment. When an inversion occurs, the chromosome is broken in two places, and the center piece flips around to produce the inversion. If either breakpoint occurs within a vital gene, the function of the gene is expected to be disrupted, possibly producing a phenotypic effect. For example, some people with hemophilia (type A) have inherited an X-linked inversion in which a breakpoint has inactivated the gene for factor VIII, a blood-clotting protein. In other cases, an inversion (or translocation) may reposition a gene on a chromosome in a way that alters its normal level of expression. This is a type of **position effect**—a change in phenotype that occurs when the position of a gene changes from one chromosomal site to a different location. This topic is also discussed in Chapter 18 (see Figure 18.2).

Because inversions seem like an unusual genetic phenomenon, it is perhaps surprising that they are found in human populations in significant numbers. About 2% of the human population carries inversions that are detectable with a light microscope. In most cases, such individuals are phenotypically normal

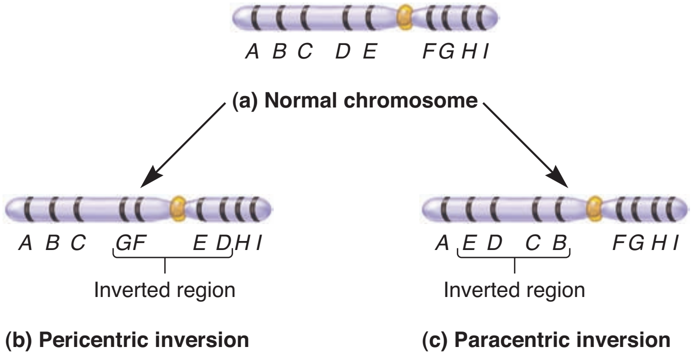


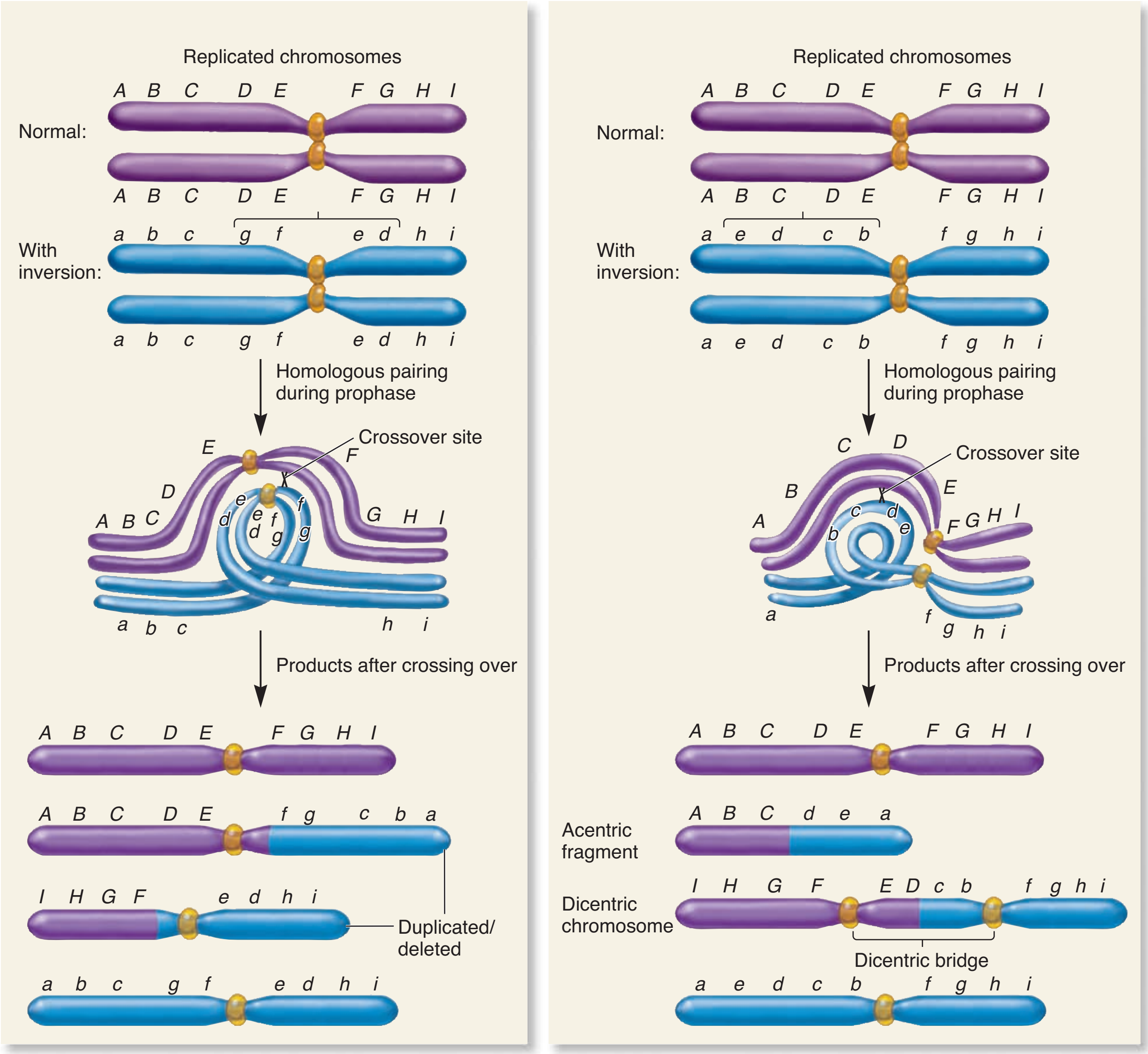
FIGURE 8.9 **Types of inversions.** **(a)** A normal chromosome with the genes ordered from *A* through *I*. A pericentric inversion **(b)** includes the centromere, whereas a paracentric inversion **(c)** does not.

and live their lives without knowing they carry an inversion. In a few cases, however, an individual with an inversion chromosome may produce offspring with phenotypic abnormalities. This event may prompt a physician to request a microscopic examination of the individual's chromosomes. In this way, phenotypically normal individuals may discover they have a chromosome with an inversion. Next, we will examine how an individual carrying an inversion may produce offspring with phenotypic abnormalities.

Inversion Heterozygotes May Produce Abnormal Chromosomes Due to Crossing Over

An individual carrying one copy of a normal chromosome and one copy of an inverted chromosome is known as an **inversion heterozygote.** Such an individual, though possibly phenotypically normal, may have a high probability of producing gametes that are abnormal in their total genetic content.

The underlying cause of gamete abnormality is the phenomenon of crossing over within the inverted region. During meiosis I, pairs of homologous sister chromatids synapse with each other. **Figure 8.10** illustrates how this occurs in an inversion heterozygote. In order for the normal chromosome and inversion chromosome to synapse properly, an **inversion loop** must form that permits the homologous genes on both chromosomes to align next to each other despite the inverted sequence. If a crossover occurs within the inversion loop, highly abnormal chromosomes are produced. A crossover is more likely to occur in this region if the inversion is large. Therefore, individuals carrying large inversions are more likely to produce abnormal gametes.


Replicated chromosomes
Normal:
A B C D E F G H I
A B C D E F G H I
With inversion:
a b c g f e d h i
a b c g f e d h i
Homologous pairing during prophase
Crossover site
E
F
D
A B C
G H I
e
d
e f
d g
f
g
a b c
h i
Products after crossing over
A B C D E F G H I
A B C D E f g c b a
I H G F e d h i
Duplicated/deleted
a b c g f e d h i
Replicated chromosomes
Normal:
A B C D E F G H I
A B C D E F G H I
With inversion:
a e d c b f g h i
a e d c b f g h i
Homologous pairing during prophase
C D
Crossover site
B
E
A
c d
b e
F G H I
a
f
g h i
Products after crossing over
A B C D E F G H I
Acentric fragment
A B C d e a
Dicentric chromosome
I H G F E D c b f g h i
Dicentric bridge
a e d c b f g h i

(a) Pericentric inversion **(b) Paracentric inversion**

FIGURE 8.10 **The consequences of crossing over in the inversion loop.** **(a)** Crossover within a pericentric inversion. **(b)** Crossover within a paracentric inversion.

Concept check: *Explain why these chromosomes can synapse only if an inversion loop forms.*

The consequences of this type of crossover depend on whether the inversion is pericentric or paracentric. Figure 8.10a describes a crossover in the inversion loop when one of the homologs has a pericentric inversion, in which the centromere lies within the inverted region of the chromosome. This event consists of a single crossover that involves only two of the four sister chromatids. Following the completion of meiosis, this single crossover yields two abnormal chromosomes. Both of these abnormal chromosomes have a segment that is deleted and a different segment that is duplicated. In this example, one of the abnormal chromosomes is missing genes *H* and *I* and has an extra copy of genes *A*, *B*, and *C*. The other abnormal chromosome has the opposite situation; it is missing genes *A*, *B*, and *C* and has an extra copy of genes *H* and *I*. These abnormal chromosomes may result in gametes that are inviable. Alternatively, if these abnormal chromosomes are passed to offspring, they are likely to produce phenotypic abnormalities, depending on the amount and nature of the duplicated and deleted genetic material. A large deletion is likely to be lethal.

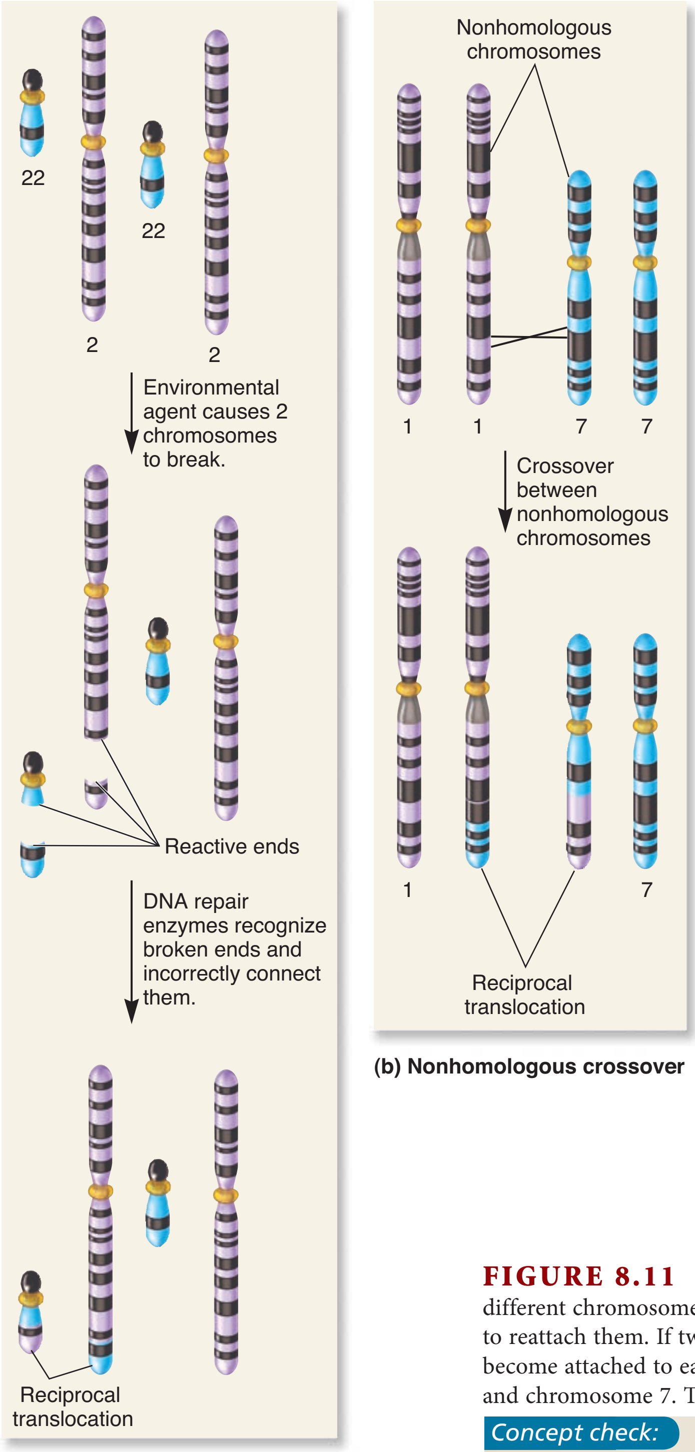
Figure 8.10b shows the outcome of a crossover involving a paracentric inversion, in which the centromere lies outside the inverted region. This single crossover event produces a very strange outcome. One chromosome, called a **dicentric** chromosome, contains two centromeres. The region of the chromosome connecting the two centromeres is called a **dicentric bridge.** The crossover also produces a piece of chromosome without any centromere—an **acentric fragment** —which is lost and degraded in subsequent cell divisions. The dicentric chromosome is a temporary condition. If the two centromeres try to move toward opposite poles during anaphase, the dicentric bridge will be forced to break at some random location. Therefore, the net result of this crossover is to produce one normal chromosome, one chromosome with an inversion, and two chromosomes that contain deletions. These two chromosomes with deletions result from the breakage of the dicentric chromosome. They are missing the genes that were located on the acentric fragment.

Translocations Involve Exchanges Between Different Chromosomes

Another type of chromosomal rearrangement is a translocation, in which a piece from one chromosome is attached to another chromosome. As described in Chapters 12 and 13, eukaryotic chromosomes have **telomeres,** which contain repeated sequences of DNA that are found at the ends of normal chromosomes. Telomeres tend to prevent translocations from occurring. They allow cells to identify where a chromosome ends and prevent the attachment of chromosomal DNA to the natural ends of a chromosome.

If cells are exposed to agents that cause chromosomes to break, the broken ends lack telomeres and are said to be reactive. A reactive end readily binds to another reactive end. If a single chromosome break occurs, DNA repair enzymes will usually recognize the two reactive ends and join them back together; the chromosome is repaired properly. However, if multiple chromosomes are broken, the reactive ends may be joined incorrectly to produce abnormal chromosomes (**Figure 8.11a**). This is one mechanism that causes reciprocal translocations to occur.

A second mechanism that can cause a translocation is a nonhomologous crossover. As shown in **Figure 8.11b**, a


22
2
22
2
Environmental agent causes 2 chromosomes to break.
Reactive ends
DNA repair enzymes recognize broken ends and incorrectly connect them.
Reciprocal translocation
Nonhomologous chromosomes
1
1
7
7
Crossover between nonhomologous chromosomes
1
7
Reciprocal translocation

(b) Nonhomologous crossover

(a) Chromosomal breakage and DNA repair

FIGURE 8.11 **Two mechanisms that cause a reciprocal translocation.** **(a)** When two different chromosomes break, the reactive ends are recognized by DNA repair enzymes, which attempt to reattach them. If two different chromosomes are broken at the same time, the incorrect ends may become attached to each other. **(b)** A nonhomologous crossover has occurred between chromosome 1 and chromosome 7. This crossover yields two chromosomes that carry translocations.

Concept check: *Which of these two mechanisms might be promoted by the presence of the same transposable element in many places in a species genome?*

reciprocal translocation can be produced when two nonhomologous chromosomes cross over. This type of rare aberrant event results in a rearrangement of the genetic material, though not a change in the total amount of genetic material.

The reciprocal translocations we have considered thus far are also called **balanced translocations** because the total amount of genetic material is not altered. Like inversions, balanced translocations usually occur without any phenotypic consequences because the individual has a normal amount of genetic material. In a few cases, balanced translocations can result in position effects similar to those that can occur in inversions. In addition, carriers of a balanced translocation are at risk of having offspring with an **unbalanced translocation,** in which significant portions of genetic material are duplicated and/or deleted. Unbalanced translocations are generally associated with phenotypic abnormalities or even lethality.

The most common type of translocation that occurs in humans is a **Robertsonian translocation,** named after William Robertson, who first described this type of fusion in grasshoppers. This type of translocation arises from breaks near the centromeres of two nonhomologous acrocentric chromosomes. In the example shown in **Figure 8.12**, the short arm of one acrocentric chromosome (chromosome 14) is exchanged with the long arm of a different acrocentric chromosome (chromosome 21). One product is a chromosome that has the long arms from both chromosomes. The other product is a very small chromosome containing the two short arms of these chromosomes; this small chromosome is usually lost during segregation. In humans, Robertsonian translocations occur at a frequency of approximately one in 900 live births. They involve only the acrocentric chromosomes 13, 14, 15, 21, and 22.

Let's consider how a person with a balanced translocation may produce gametes and offspring with an unbalanced translocation. An inherited human syndrome known as familial Down syndrome provides an example. A person with a normal phenotype may have one copy of chromosome 14, one copy of chromosome 21, and one copy of a Robertsonian translocation that contains the long arms of chromosomes 14 and 21 (**Figure 8.13a**). The individual has a normal phenotype because the total amount of genetic material is present (with the exception of the short arms of these chromosomes that do not carry vital genetic material). During meiosis, these three types of chromosomes replicate and segregate from each other. However, because the three chromosomes cannot segregate evenly, six possible types of gametes may be produced. One gamete is normal, and one is a balanced carrier of a translocated chromosome. The four gametes to the right, however, are unbalanced, either containing too much or too little material from chromosome 14 or 21. The unbalanced gametes may be inviable, or they could combine with a normal gamete. The resulting three offspring on the right will not survive. In comparison, the unbalanced gamete that carries chromosome 21 and the fused chromosome results in an offspring with familial Down syndrome (also see karyotype in **Figure 8.13b**). Such an offspring has three copies of the genes that are found on the long arm of chromosome 21. **Figure 8.13c** shows a person with this disorder. She has characteristics similar to those of an individual who has the more prevalent form of Down syndrome, which is due to three entire copies of chromosome 21. We will examine this common form of Down syndrome later in this chapter.

Individuals with Reciprocal Translocations May Produce Abnormal Gametes Due to the Pairing and Segregation of Chromosomes

As we have seen, individuals who carry balanced translocations have a greater risk of producing gametes with unbalanced combinations of chromosomes. Whether or not this occurs depends on the segregation pattern during meiosis I (**Figure 8.14**). In this example, the parent carries a reciprocal translocation and is likely to be phenotypically normal. During meiosis, the homologous chromosomes attempt to synapse with each other. Because of the translocations, the pairing of homologous regions leads to the formation of an unusual structure that contains four pairs of sister chromatids (i.e., eight chromatids), termed a **translocation cross.**

To understand the segregation of translocated chromosomes, pay close attention to the centromeres, which are numbered in Figure 8.14. For these translocated chromosomes, the expected segregation pattern is governed by the centromeres. Each haploid gamete should receive one centromere located on chromosome 1 and one centromere located on chromosome 2. This can occur in two ways. One possibility is alternate segregation. As shown in Figure 8.14a, this occurs when the chromosomes diagonal to each other within the translocation cross sort into the same cell. One daughter cell receives two normal chromosomes, and the other cell gets two translocated chromosomes. Following meiosis II, four haploid cells are produced: two have normal chromosomes, and two have reciprocal (balanced) translocations.

Another possible segregation pattern is called adjacent-1 segregation (Figure 8.14b). This occurs when adjacent chromosomes (one of each type of centromere) segregate into the same cell. Following anaphase of meiosis I, each daughter cell receives one normal chromosome and one translocated chromosome. After meiosis II is completed, four haploid cells are produced, all

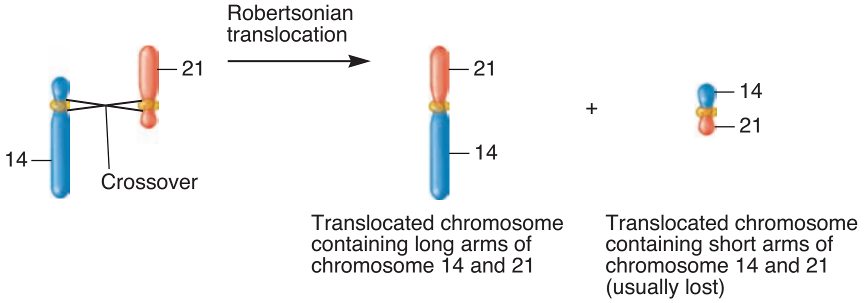
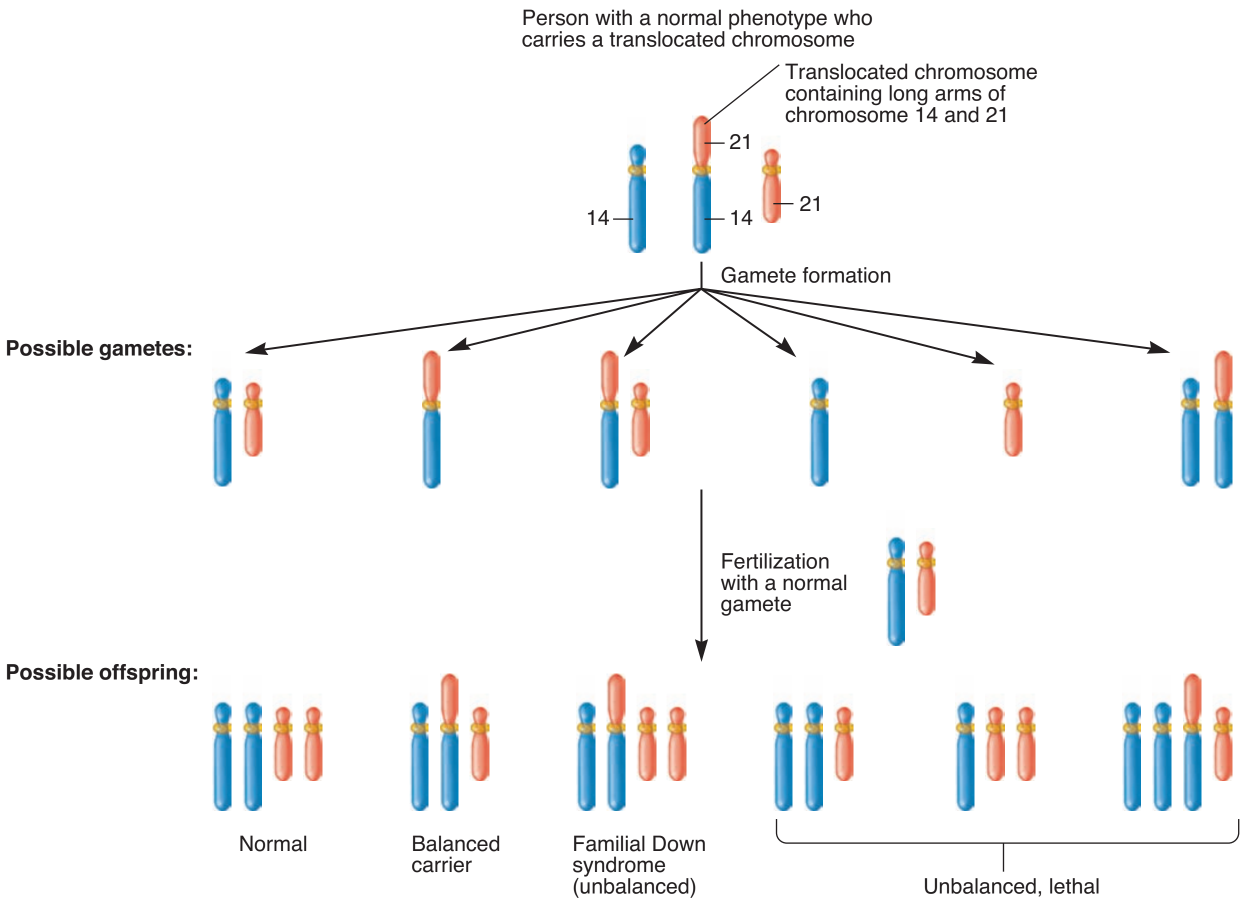
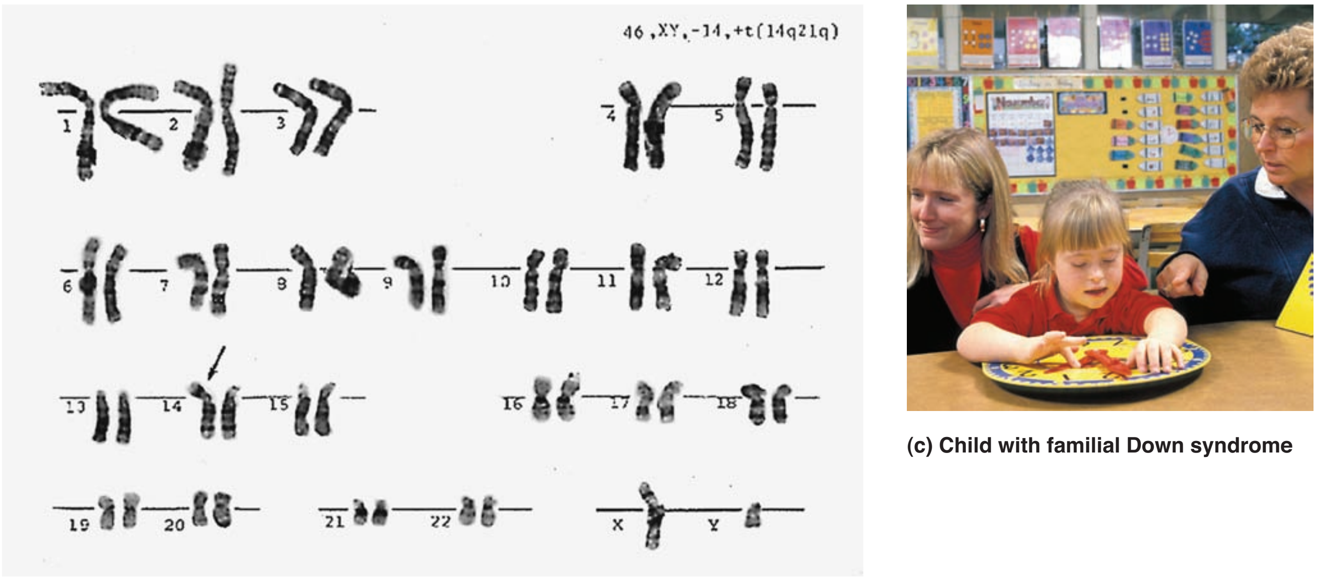

Robertsonian translocation
21
14
Crossover
21
14
+
14
21
Translocated chromosome containing long arms of chromosome 14 and 21
Translocated chromosome containing short arms of chromosome 14 and 21 (usually lost)

FIGURE 8.12 **Robertsonian translocation.** In this example, the long arm of chromosome 21 is exchanged with the short arm of chromosome 14.


Person with a normal phenotype who carries a translocated chromosome
Translocated chromosome containing long arms of chromosome 14 and 21
21
14
14
21
Gamete formation
Possible gametes:
Fertilization with a normal gamete
Possible offspring:
Normal
Balanced carrier
Familial Down syndrome (unbalanced)
Unbalanced, lethal

(a) Possible transmission patterns


46,XY,-14,+t(14q21q)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

(c) Child with familial Down syndrome

(b) Karyotype of a male with familial Down syndrome

FIGURE 8.13 **Transmission of familial Down syndrome.** **(a)** Potential transmission of familial Down syndrome. The individual with the chromosome composition shown at the top of this figure may produce a gamete carrying chromosome 21 and a fused chromosome containing the long arms of chromosome 14 and 21. Such a gamete can give rise to an offspring with familial Down syndrome. **(b)** The karyotype of an individual with familial Down syndrome. This karyotype shows that the long arm of chromosome 21 has been translocated to chromosome 14 (see arrow). In addition, the individual also carries two normal copies of chromosome 21. **(c)** An individual with this disorder.

Concept check: If these segregation patterns are equally likely, what is the probability that a gamete made from this individual will result in a viable offspring with a normal phenotype?

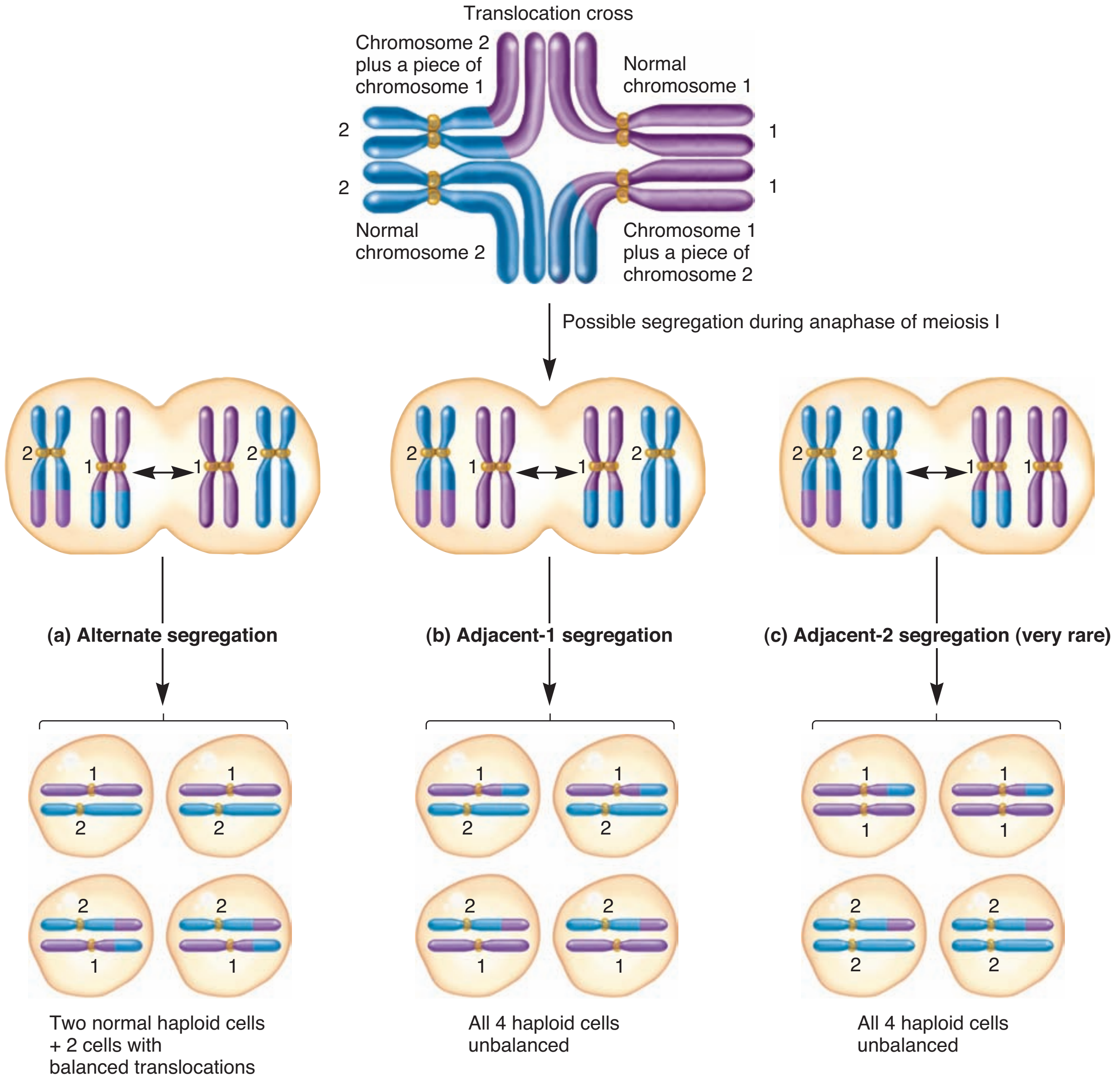

Translocation cross
Chromosome 2 plus a piece of chromosome 1
Normal chromosome 1
2
2
1
1
Normal chromosome 2
Chromosome 1 plus a piece of chromosome 2
Possible segregation during anaphase of meiosis I
2
1
1
2
(a) Alternate segregation
(b) Adjacent-1 segregation
(c) Adjacent-2 segregation (very rare)
Two normal haploid cells + 2 cells with balanced translocations
All 4 haploid cells unbalanced
All 4 haploid cells unbalanced

FIGURE 8.14 **Meiotic segregation of a reciprocal translocation.** Follow the numbered centromeres through each process. **(a)** Alternate segregation gives rise to balanced haploid cells, whereas **(b)** adjacent-1 and **(c)** adjacent-2 segregation produce haploid cells with an unbalanced amount of genetic material.

Concept check: *Explain why these chromosomes form a translocation cross during prophase of meiosis I.*

of which are genetically unbalanced, because part of one chromosome has been deleted and part of another has been duplicated. If these haploid cells give rise to gametes that unite with a normal gamete, the zygote is expected to be abnormal genetically and possibly phenotypically.

On very rare occasions, adjacent-2 segregation can occur (Figure 8.14c). In this case, the centromeres do not segregate as they should. One daughter cell has received both copies of the centromere on chromosome 1; the other, both copies of the centromere on chromosome 2. This rare segregation pattern also yields four abnormal haploid cells that contain an unbalanced combination of chromosomes.

Alternate and adjacent-1 segregation patterns are the likely outcomes when an individual carries a reciprocal translocation. Depending on the sizes of the translocated segments, both types may be equally likely to occur. In many cases, the haploid cells from adjacent-1 segregation are not viable, thereby lowering the fertility of the parent. This condition is called **semisterility.**

8.4 REVIEWING THE KEY CONCEPTS

- Inversions can be pericentric or paracentric. In an inversion heterozygote, crossing over within the inversion loop can create deletions and duplications in the resulting chromosomes (see Figures 8.9, 8.10).
- Two mechanisms that may produce translocations are chromosome breakage and subsequent repair, and nonhomologous crossing over (see Figure 8.11).

- A Robertsonian translocation results in a chromosome in which the long arms of two different acrocentric chromosomes become joined (see Figure 8.12).
- Familial Down syndrome is due to a translocation between chromosome 14 and 21 (see Figure 8.13).
- Individuals that carry a balanced translocation may have a high probability of producing unbalanced gametes, depending on the segregation pattern during meiosis I (see Figure 8.14).

8.4 COMPREHENSION QUESTIONS

1. A paracentric inversion
   a. includes the centromere within the inverted region.
   b. does not include the centromere within the inverted region.
   c. has two adjacent inverted regions.
   d. has an inverted region at the very end of a chromosome.
2. Due to crossing over within an inversion loop, a heterozygote with a pericentric inversion may produce gametes that
   a. carry a deletion.
   b. carry a duplication.
   c. carry a translocation.
   d. both a and b.
3. A mechanism that may cause a translocation is
   a. the joining of reactive ends when two different chromosomes break.
   b. crossing over between nonhomologous chromosomes.
   c. crossing over between homologous chromosomes.
   d. both a and b.

8.5 CHANGES IN CHROMOSOME NUMBER: AN OVERVIEW

As we saw in Sections 8.1 to 8.4, chromosome structure can be altered in a variety of ways. Likewise, the total number of chromosomes can vary. Eukaryotic species typically contain several chromosomes that are inherited as one or more sets. Variations in chromosome number can be categorized in two ways: variation in the number of sets of chromosomes and variation in the number of particular chromosomes within a set.

Organisms that are **euploid** have a chromosome number that is an exact multiple of a chromosome set. In *Drosophila melanogaster,* for example, a normal individual has 8 chromosomes. The species is diploid, having two sets of 4 chromosomes each (**Figure 8.15a**). A normal fruit fly is euploid because 8 chromosomes divided by 4 chromosomes per set equals two exact sets. On rare occasions, an abnormal fruit fly can be produced with 12 chromosomes, containing three sets of 4 chromosomes each. This alteration in euploidy produces a **triploid** fruit fly with 12

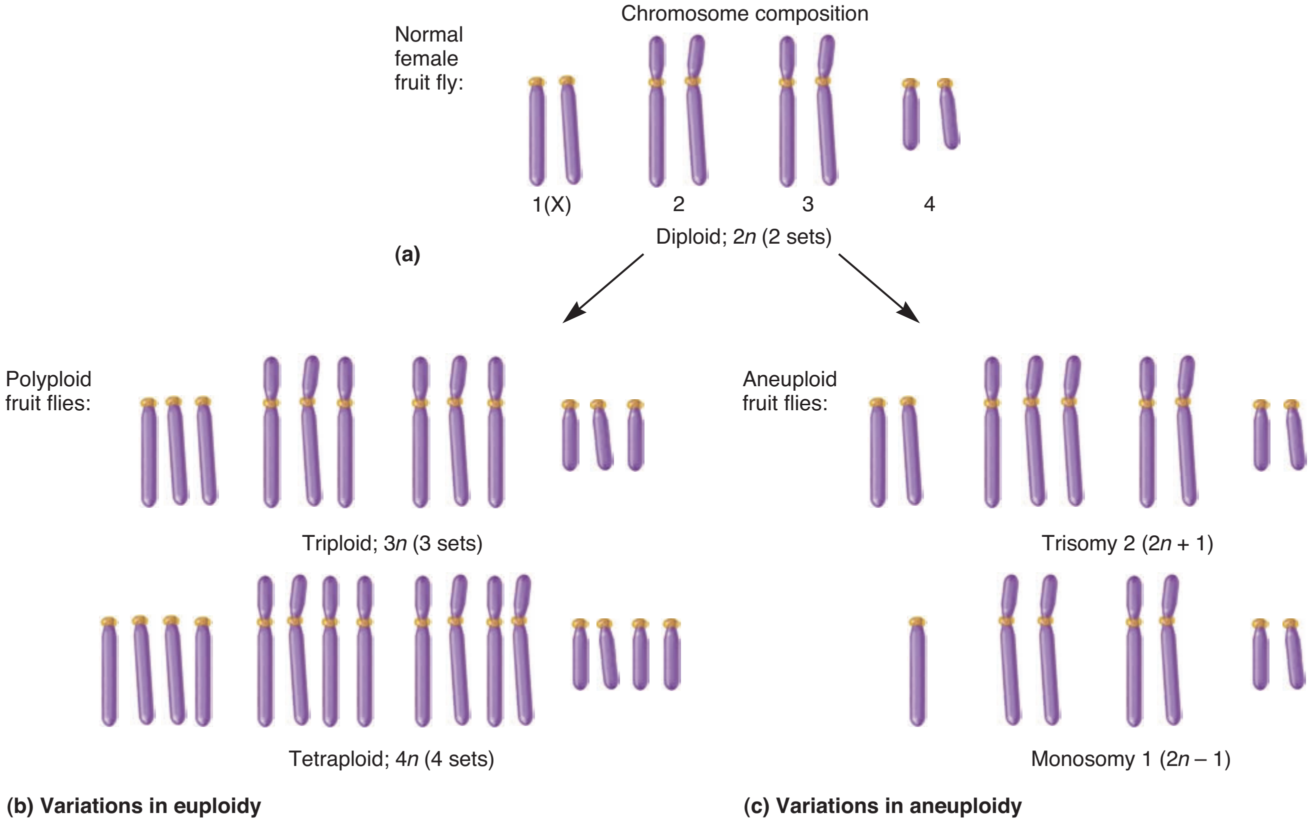


FIGURE 8.15 **Types of variation in chromosome number.** **(a)** The normal diploid number of chromosomes in *Drosophila.* **(b)** Examples of polyploidy. **(c)** Examples of aneuploidy.

Concept check: What adjectives can be used to describe a fruit fly that has a total of seven chromosomes because it is missing one copy of chromosome 3?

chromosomes. Such a fly is also euploid because it has exactly three sets of chromosomes. Organisms with three or more sets of chromosomes are also called **polyploid** (**Figure 8.15b**). Geneticists use the letter *n* to represent a set of chromosomes. A diploid organism is referred to as $2n$, a triploid organism as $3n$, a **tetraploid** organism as $4n$, and so on.

A second way in which chromosome number can vary is by **aneuploidy.** Such variation involves an alteration in the number of particular chromosomes, so the total number of chromosomes is not an exact multiple of a set. For example, an abnormal fruit fly could contain nine chromosomes instead of eight because it has three copies of chromosome 2 instead of the normal two copies (**Figure 8.15c**). Such an animal is said to have trisomy 2 or to be **trisomic.** Instead of being perfectly diploid ($2n$), a trisomic animal is $2n + 1$. By comparison, a fruit fly could be lacking a single chromosome, such as chromosome 1, and contain a total of seven chromosomes ($2n - 1$). This animal is **monosomic** and is described as having monosomy 1.

8.5 REVIEWING THE KEY CONCEPTS

- Variation in the number of chromosomes may involve changes in the number of sets (euploidy) or changes in the number of particular chromosomes within a set (aneuploidy) (see Figure 8.15).

8.5 COMPREHENSION QUESTIONS

1. Humans have 23 chromosomes per set. A person with 45 chromosomes could be described as being
   a. euploid.
   b. aneuploid.
   c. monoploid.
   d. trisomic.

8.6 VARIATION IN THE NUMBER OF CHROMOSOMES WITHIN A SET: ANEUPLOIDY

In this section, we will consider several examples of aneuploidy. As you will learn, this is generally regarded as an abnormal condition that usually has a negative effect on phenotype.

Aneuploidy Causes an Imbalance in Gene Expression That Is Often Detrimental to the Phenotype of the Individual

The phenotype of every eukaryotic species is influenced by thousands of different genes. In humans, for example, a single set of chromosomes contains approximately 20,000 to 25,000 different genes. To produce a phenotypically normal individual, intricate coordination has to occur in the expression of thousands of genes. In the case of humans and other diploid species, evolution has resulted in a developmental process that works correctly when somatic cells have two copies of each chromosome. In other words, when a human is diploid, the balance of gene expression among many different genes usually produces a person with a normal phenotype.

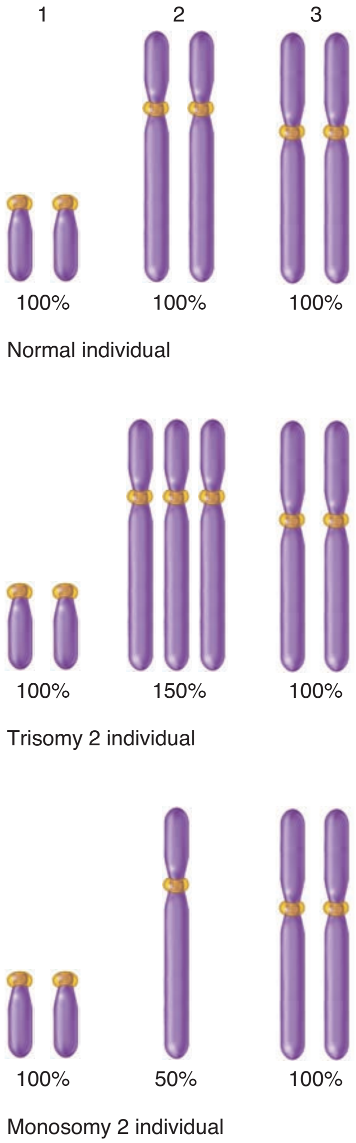


FIGURE 8.16 **Imbalance of gene products in trisomic and monosomic individuals.** Aneuploidy of chromosome 2 (i.e., trisomy and monosomy) leads to an imbalance in the amount of gene products from chromosome 2 compared with the amounts from chromosomes 1 and 3.

Concept check: Describe the imbalance in gene products that would occur in monosomy 2.

Aneuploidy commonly causes an abnormal phenotype. To understand why, let's consider the relationship between gene expression and chromosome number in a species that has three pairs of chromosomes (**Figure 8.16**). The level of gene expression is influenced by the number of genes per cell. Compared with a diploid cell, if a gene is carried on a chromosome that is present in three copies instead of two, more of the gene product is typically made. For example, a gene present in three copies instead of two may produce 150% of the gene product, though that number may vary due to effects of gene regulation. Alternatively, if only one copy of that gene is present due to a missing chromosome, less of the gene product is usually made, perhaps only 50%. Therefore, in trisomic and monosomic individuals, an imbalance occurs between the level of gene expression on the chromosomes found in pairs versus the expression level on the chromosomes that are not.

At first glance, the difference in gene expression between euploid and aneuploid individuals may not seem terribly dramatic. Keep in mind, however, that a eukaryotic chromosome carries hundreds or even thousands of different genes. Therefore, when an organism is trisomic or monosomic, many gene products occur in excessive or deficient amounts. This imbalance among many genes appears to underlie the abnormal phenotypic effects that aneuploidy frequently causes. In most cases, these effects are detrimental and produce an individual that is less likely to survive than a euploid individual.

Aneuploidy in Humans Causes Abnormal Phenotypes

A key reason why geneticists are so interested in aneuploidy is its relationship to certain inherited disorders in humans. Even though most people are born with a normal number of chromosomes (i.e., 46), alterations in chromosome number occur fairly frequently during gamete formation. About 5% to 10% of all fertilized human eggs result in an embryo with an abnormality in chromosome number! In most cases, these abnormal embryos do not develop properly and result in a spontaneous abortion very early in pregnancy. Approximately 50% of all spontaneous abortions are due to alterations in chromosome number.

In some cases, an abnormality in chromosome number produces an offspring that survives to birth or longer. Several human disorders involve abnormalities in chromosome number. The most common are trisomies of chromosomes 13, 18, or 21, and abnormalities in the number of the sex chromosomes (**Table 8.1**). Most of the known trisomies involve chromosomes that are relatively small and carry fewer genes compared to larger chromosomes. Trisomies of the other human autosomes and monosomies of all autosomes are presumed to produce a lethal phenotype, and many have been found in spontaneously aborted embryos and fetuses. For example, all possible human trisomies have been found in spontaneously aborted embryos except trisomy 1. It is believed that trisomy 1 is lethal at such an early stage that it prevents the successful implantation of the embryo.

Variation in the number of X chromosomes, unlike that of other large chromosomes, is often nonlethal. The survival of trisomy X individuals may be explained by X inactivation, which is described in Chapter 4. In an individual with more than one X chromosome, all additional X chromosomes are converted to Barr bodies in the somatic cells of adult tissues. In an individual with trisomy X, for example, two out of three X chromosomes are converted to inactive Barr bodies. Unlike the level of expression for autosomal genes, the normal level of expression for X-linked genes is from a single X chromosome. In other words, the correct level of mammalian gene expression results from two copies of each autosomal gene and one copy of each X-linked gene. This explains how the expression of X-linked genes in males (XY) can be maintained at the same levels as in females (XX). It may also explain why trisomy X is not a lethal condition. The phenotypic effects noted in Table 8.1 involving sex chromosomal abnormalities may be due to the expression of X-linked genes prior to embryonic X inactivation or to the expression of genes on the inactivated X chromosome. As described in Chapter 4, pseudoautosomal genes and some other genes on the inactivated X chromosome are expressed in humans. Having one or three copies of the sex chromosomes results in an under- or overexpression of these X-linked genes, respectively.

Human abnormalities in chromosome number are influenced by the age of the parents. Older parents are more likely to produce children with abnormalities in chromosome number. Down syndrome provides an example. This syndrome was first described by the English physician John Langdon Down in 1866. It occurs at a frequency of about 1 in 800 live births. The symptoms are described in Table 8.1. The common form of this disorder is caused by the inheritance of three copies of chromosome 21. Therefore, Down syndrome is also known as trisomy 21. The incidence of Down syndrome rises with the age of either parent. In males, however, the rise occurs relatively late in life, usually past the age when most men have children. By comparison, the likelihood of having a child with Down syndrome rises dramatically during a woman's reproductive age (**Figure 8.17**). The association between maternal age and Down syndrome was discovered by L. S. Penrose in 1933, even before the chromosomal basis for the disorder was identified by the French scientist Jérôme Lejeune in 1959. Down syndrome is most commonly caused by **nondisjunction,** which means that the chromosomes do not segregate properly. (Nondisjunction is discussed later in

TABLE **8.1**

Aneuploid Conditions in Humans

Condition	Frequency	Syndrome	Characteristics
Autosomal			
Trisomy 13	1/15,000	Patau	Mental and physical deficiencies, wide variety of defects in organs, large triangular nose, early death
Trisomy 18	1/6000	Edward	Mental and physical deficiencies, facial abnormalities, extreme muscle tone, early death
Trisomy 21	1/800	Down	Mental deficiencies, abnormal pattern of palm creases, slanted eyes, flattened face, short stature
Sex Chromosomal			
XXY	1/1000	Klinefelter	Sexual immaturity (no sperm), breast swelling (males)
XYY	1/1000	Jacobs	Tall and thin (males)
XXX	1/1500	Triple X	Tall and thin, menstrual irregularity (females)
X0	1/5000	Turner	Short stature, webbed neck, sexually undeveloped (females)

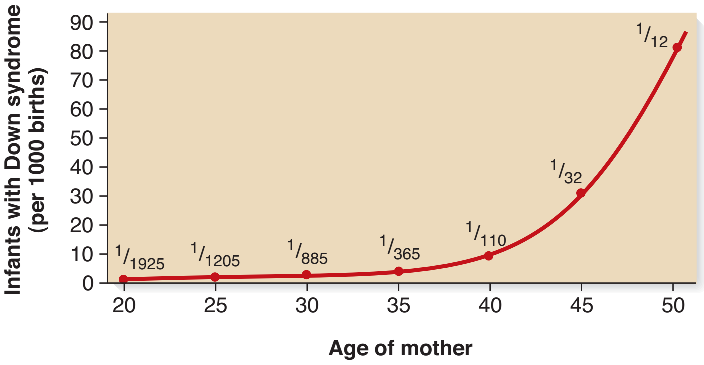


FIGURE 8.17 **The incidence of Down syndrome births according to the age of the mother.** The *y*-axis shows the number of infants born with Down syndrome per 1000 live births, and the *x*-axis plots the age of the mother at the time of birth. The data points indicate the fraction of live offspring born with Down syndrome.

this chapter.) In this case, nondisjunction of chromosome 21 usually occurs during meiosis I in the oocyte.

Different hypotheses have been proposed to explain the relationship between maternal age and Down syndrome. One popular idea suggests that it may be due to the age of the oocytes. Human primary oocytes are produced within the ovary of the female fetus prior to birth and are arrested at prophase of meiosis I and remain in this stage until the time of ovulation. Therefore, as a woman ages, her primary oocytes have been in prophase I for a progressively longer period of time. This added length of time may contribute to an increased frequency of nondisjunction, though direct evidence to support this claim is lacking. About 5% of the time, Down syndrome is due to an extra paternal chromosome. Prenatal tests can determine if a fetus has Down syndrome and some other genetic abnormalities. The topic of genetic testing is discussed in Chapter 23.

8.6 REVIEWING THE KEY CONCEPTS

- Aneuploidy is often detrimental due to an imbalance in gene expression. Down syndrome, an example of an aneuploidy in humans, increases in frequency with maternal age (see Figures 8.16, 8.17, Table 8.1).

8.6 COMPREHENSION QUESTIONS

1. In a trisomic individual, such as a person with trisomy 21 (Down syndrome), a genetic imbalance occurs because
   a. genes on chromosome 21 are overexpressed.
   b. genes on chromosome 21 are underexpressed.
   c. genes on the other chromosomes are overexpressed.
   d. genes on the other chromosomes are underexpressed.
2. Humans that survive with aneuploidy usually have incorrect numbers of chromosome 13, 18, or 21 or the sex chromosomes. A possible explanation why these abnormalities permit survival is because
   a. the chromosomes have clusters of genes that aid in embryonic growth.
   b. the chromosomes are small and carry relatively few genes.
   c. X inactivation.
   d. both b and c.

8.7 VARIATION IN THE NUMBER OF SETS OF CHROMOSOMES

We now turn our attention to changes in the number of sets of chromosomes, referred to as variations in euploidy. In some cases, such changes are detrimental. However, in many species, particularly plants, additional sets of chromosomes are very common and are often beneficial with regard to an individual's phenotype.

Variations in Euploidy Occur Naturally in a Few Animal Species

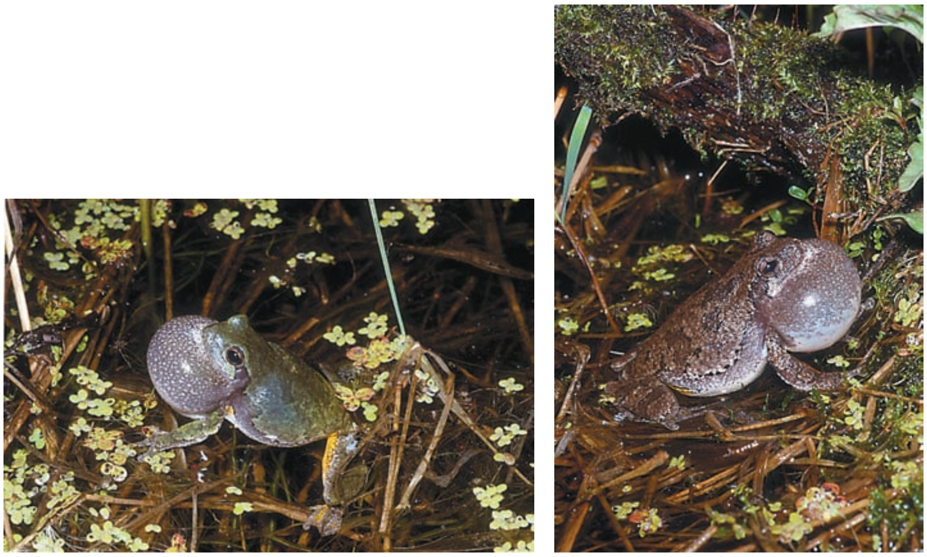
Most species of animals are diploid. In some cases, changes in euploidy are not well tolerated. For example, polyploidy in mammals is often a lethal condition. However, many examples of naturally occurring variations in euploidy occur. In **haplodiploid** species, which includes many species of bees, wasps, and ants, one of the sexes is haploid, usually the male, and the other is diploid (see Chapter 4, Figure 4.2). For example, male bees, which are called drones, contain a single set of chromosomes. They are produced from unfertilized eggs. By comparison, female bees are produced from fertilized eggs and are diploid.

Many examples of vertebrate polyploid animals have been discovered. Interestingly, on several occasions, animals that are morphologically very similar can be found as a diploid species as well as a separate polyploid species. This situation occurs among certain amphibians and reptiles. **Figure 8.18** shows photographs of a diploid and a tetraploid (4*n*) frog. As you can see, they look nearly indistinguishable from each other. Their difference can be revealed only by an examination of the chromosome number in the somatic cells of the animals and by mating calls—*H. chrysoscelis* has a faster trill rate than *H. versicolor*.

Variations in Euploidy Can Occur in Certain Tissues Within an Animal

Thus far, we have considered variations in chromosome number that occur at fertilization, so all somatic cells of an individual contain this variation. In many animals, certain tissues of the body display normal variations in the number of sets of chromosomes. Diploid animals sometimes produce tissues that are polyploid. For example, the cells of the human liver can vary to a great degree in their ploidy. Liver cells contain nuclei that can be triploid, tetraploid, and even octaploid (8*n*). The occurrence of polyploid tissues or cells in organisms that are otherwise diploid is known as **endopolyploidy.** What is the biological significance of endopolyploidy? One possibility is that the increase in chromosome number in certain cells may enhance their ability to produce specific gene products that are needed in great abundance.

An unusual example of natural variation in the ploidy of somatic cells occurs in *Drosophila* and some other insects. Within certain tissues, such as the salivary glands, the chromosomes undergo repeated rounds of chromosome replication without cellular division. For example, in the salivary gland cells of *Drosophila,* the pairs of chromosomes double approximately nine



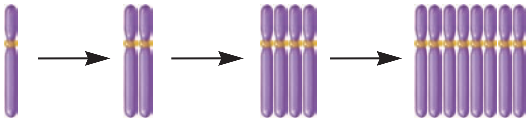
(a) ***Hyla chrysoscelis*** **(b)** ***Hyla versicolor***

FIGURE 8.18 Differences in euploidy in two closely related frog species. The frog in **(a)** is diploid, whereas the frog in **(b)** is tetraploid.

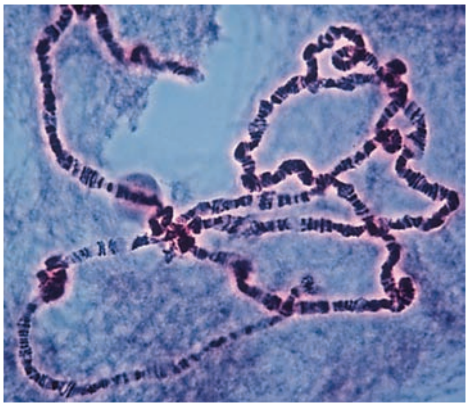
Genes → Traits Though similar in appearance, these two species differ in their number of chromosome sets. At the level of gene expression, this observation suggests that the number of copies of each gene (two versus four) does not critically affect the phenotype of these two species.

times ($2^9 = 512$). **Figure 8.19a** illustrates how repeated rounds of chromosomal replication produce a bundle of chromosomes that lie together in a parallel fashion. This bundle, termed a **polytene chromosome,** was first observed by E. G. Balbiani in 1881. Later, in the 1930s, Theophilus Painter and colleagues recognized that the size and morphology of polytene chromosomes provided geneticists with unique opportunities to study chromosome structure and gene organization.

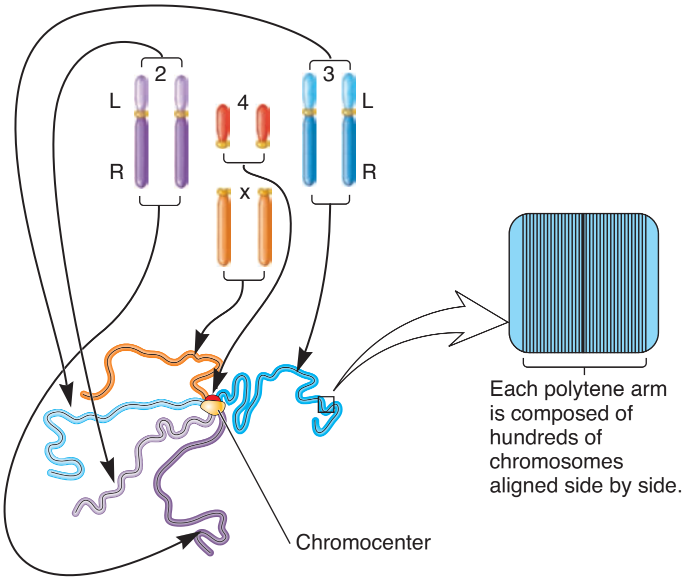
Figure 8.19b shows a micrograph of a polytene chromosome. The structure of polytene chromosomes is different from other forms of endopolyploidy because the replicated chromosomes remain attached to each other. Prior to the formation of



(a) Repeated chromosome replication produces polytene chromosome.



(b) A polytene chromosome


2
L
R
4
X
3
L
R
Each polytene arm is composed of hundreds of chromosomes aligned side by side.
Chromocenter

(c) Relationship between a polytene chromosome and regular *Drosophila* chromosomes

FIGURE 8.19 Polytene chromosomes in *Drosophila*. (a) A schematic illustration of the formation of polytene chromosomes. Several rounds of repeated replication without cellular division result in a bundle of sister chromatids that lie side by side. Both homologs also lie parallel to each other. This replication does not occur in highly condensed, heterochromatic DNA near the centromere. **(b)** A photograph of a polytene chromosome. **(c)** This drawing shows the relationship between the four pairs of chromosomes and the formation of a polytene chromosome in the salivary gland. The heterochromatic regions of the chromosomes aggregate at the chromocenter, and the arms of the chromosomes project outward. In chromosomes with two arms, the short arm is labeled L and the long arm is labeled R.

Concept check: *Approximately how many copies of chromosome 2 are found in a polytene chromosome?*

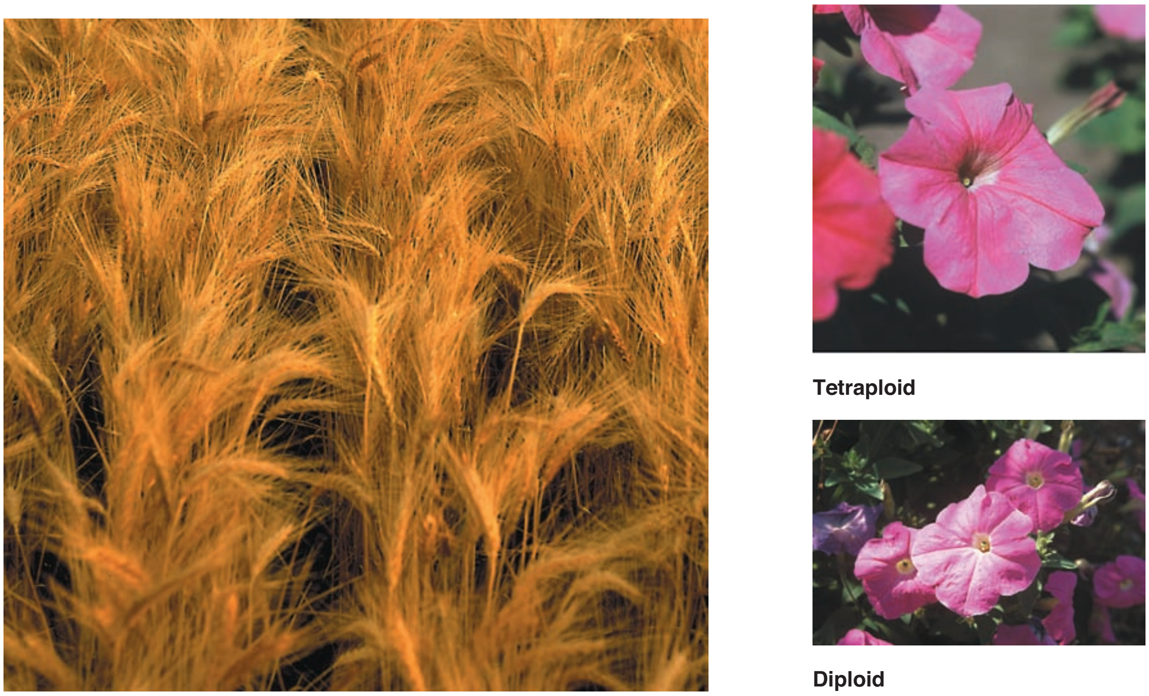
polytene chromosomes, *Drosophila* cells contain eight chromosomes (two sets of four chromosomes each; see Figure 8.15a). In the salivary gland cells, the homologous chromosomes synapse with each other and replicate to form a polytene structure. During this process, the four types of chromosomes aggregate to form a single structure with several polytene arms. The central point where the chromosomes aggregate is known as the **chromocenter.** Each of the four types of chromosome is attached to the chromocenter near its centromere. The X and Y and chromosome 4 are telocentric, and chromosomes 2 and 3 are metacentric. Therefore, chromosomes 2 and 3 have two arms that radiate from the chromocenter, whereas the X and Y and chromosome 4 have a single arm projecting from the chromocenter (**Figure 8.19c**).

Variations in Euploidy Are Common in Plants

We now turn our attention to variations of euploidy that occur in plants. Compared with animals, plants more commonly exhibit polyploidy. Among ferns and flowering plants, at least 30% to 35% of species are polyploid. Polyploidy is also important in agriculture. Many of the fruits and grains we eat are produced from polyploid plants. For example, the species of wheat that we use to make bread, *Triticum aestivum,* is a hexaploid (6*n*) that arose from the union of diploid genomes from three closely related species (**Figure 8.20a**). Different species of strawberries are diploid, tetraploid, hexaploid, and even octaploid!

In many instances, polyploid strains of plants display outstanding agricultural characteristics. They are often larger in size and more robust. These traits are clearly advantageous in the production of food. In addition, polyploid plants tend to exhibit a greater adaptability, which allows them to withstand harsher environmental conditions. Also, polyploid ornamental plants often produce larger flowers than their diploid counterparts (**Figure 8.20b**).

Polyploid plants having an odd number of chromosome sets, such as triploids (3*n*) or pentaploids (5*n*), usually cannot reproduce. Why are they sterile? The sterility arises because they produce highly aneuploid gametes, which are nonviable. During prophase of meiosis I, homologous pairs of sister chromatids form bivalents. However, organisms with an odd number of chromosomes, such as three, display an unequal separation of homologous chromosomes during anaphase of meiosis I (**Figure 8.21**). An odd number cannot be divided equally between two daughter cells. For each type of chromosome, a daughter cell randomly gets one or two copies. For example, one daughter cell might receive one copy of chromosome 1, two copies of chromosome 2, two copies of chromosome 3, one copy of chromosome 4, and so forth. For a



(a) Cultivated wheat, a hexaploid species

(b) A comparison of diploid and tetraploid petunias

FIGURE 8.20 **Examples of polyploid plants.** **(a)** Cultivated wheat, *Triticum aestivum,* is a hexaploid. It was derived from three different diploid species of grasses that originally were found in the Middle East and were cultivated by ancient farmers in that region. **(b)** Differences in euploidy may exist in two closely related petunia species. The flower at the bottom is diploid, whereas the large one at the top is tetraploid.

Genes → Traits An increase in chromosome number from diploid to tetraploid or hexaploid affects the phenotype of the individual. In the case of many plant species, a polyploid individual is larger and more robust than its diploid counterpart. This suggests that having additional copies of each gene is somewhat better than having two copies of each gene. This phenomenon in plants is rather different from the situation in animals. Tetraploidy in animals may have little effect (as in Figure 8.18b), but it is also common for polyploidy in animals to be detrimental.

Concept check: *What are some common advantages of polyploidy in plants?*

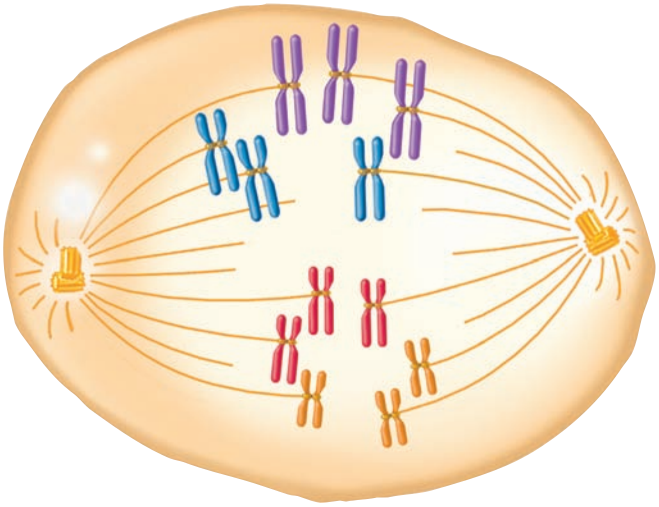


FIGURE 8.21 **Schematic representation of anaphase of meiosis I in a triploid organism containing three sets of four chromosomes.** In this example, the homologous chromosomes (three each) do not evenly separate during anaphase. Each cell receives one copy of some chromosomes and two copies of other chromosomes. This produces aneuploid gametes.

Concept check: Explain why a triploid individual is usually infertile.

triploid species containing many different chromosomes in a set, meiosis is very unlikely to produce a daughter cell that is euploid. If we assume that a daughter cell receives either one copy or two copies of each kind of chromosome, the probability that meiosis will produce a cell that is perfectly haploid or diploid is $(1/2)^{n-1}$, where n is the number of chromosomes in a set. As an example, in a triploid organism containing 20 chromosomes per set, the probability of producing a haploid or diploid cell is 0.000001907, or 1 in 524,288. Thus, meiosis is almost certain to produce cells that contain one copy of some chromosomes and two copies of the others. This high probability of aneuploidy underlies the reason for triploid sterility.

Though sterility is generally a detrimental trait, it can be desirable agriculturally because it may result in a seedless fruit. For example, domestic bananas and seedless watermelons are triploid varieties. The domestic banana was originally derived from a seed-producing diploid species and has been asexually propagated by humans via cuttings. The small black spots in the center of a domestic banana are degenerate seeds. In the case of flowers, the seedless phenotype can also be beneficial. Seed producers such as Burpee have developed triploid varieties of flowering plants such as marigolds. Because the triploid marigolds are sterile and unable to set seed, more of their energy goes into flower production. According to Burpee, "They bloom and bloom, unweakened by seed bearing."

8.7 REVIEWING THE KEY CONCEPTS

- Among animals, variation in euploidy is relatively rare, though it does occur. Some tissues within an animal may exhibit polyploidy. An example is polytene chromosomes found in salivary cells of *Drosophila* (see Figures 8.18, 8.19).
- Polyploidy in plants is relatively common and has found many advantages in agriculture. Triploid plants are usually seedless because they cannot segregate their chromosomes evenly during meiosis (see Figures 8.20, 8.21).

8.7 COMPREHENSION QUESTIONS

1. The term *endopolyploidy* refers to the phenomenon of having
   a. too many chromosomes.
   b. extra chromosomes inside the cell nucleus.
   c. certain cells of the body with extra sets of chromosomes.
   d. extra sets of chromosomes in gametes.
2. In agriculture, an advantage of triploid plants is they are
   a. more fertile.
   b. often seedless.
   c. always disease-resistant.
   d. all of the above.

8.8 MECHANISMS THAT PRODUCE VARIATION IN CHROMOSOME NUMBER

As we have seen, variations in chromosome number are fairly widespread and usually have a significant effect on the phenotypes of plants and animals. For these reasons, researchers have wanted to understand the cellular mechanisms that cause variations in chromosome number. In some cases, a change in chromosome number is the result of nondisjunction. The term *nondisjunction* refers to an event in which the chromosomes do not segregate properly. As we will see, it may be caused by an improper separation of homologous pairs in a bivalent in meiosis or a failure of the centromeres to disconnect during mitosis.

Meiotic nondisjunction can produce cells that have too many or too few chromosomes. If such a cell gives rise to a gamete that fuses with a normal gamete during fertilization, the resulting offspring will have an abnormal chromosome number in all of its cells. An abnormal nondisjunction event also may occur after fertilization in one of the somatic cells of the body. This second mechanism is known as **mitotic nondisjunction.** When this occurs during embryonic stages of development, it may lead to a patch of tissue in the organism that has an altered chromosome number. A third common way in which the chromosome number of an organism can vary is by interspecies crosses. An **alloploid** organism contains sets of chromosomes from two or more different species. This term refers to the occurrence of chromosome sets (ploidy) from the genomes of different (allo) species. In this section, we will examine these three mechanisms in greater detail.

Meiotic Nondisjunction Can Produce Aneuploidy or Polyploidy

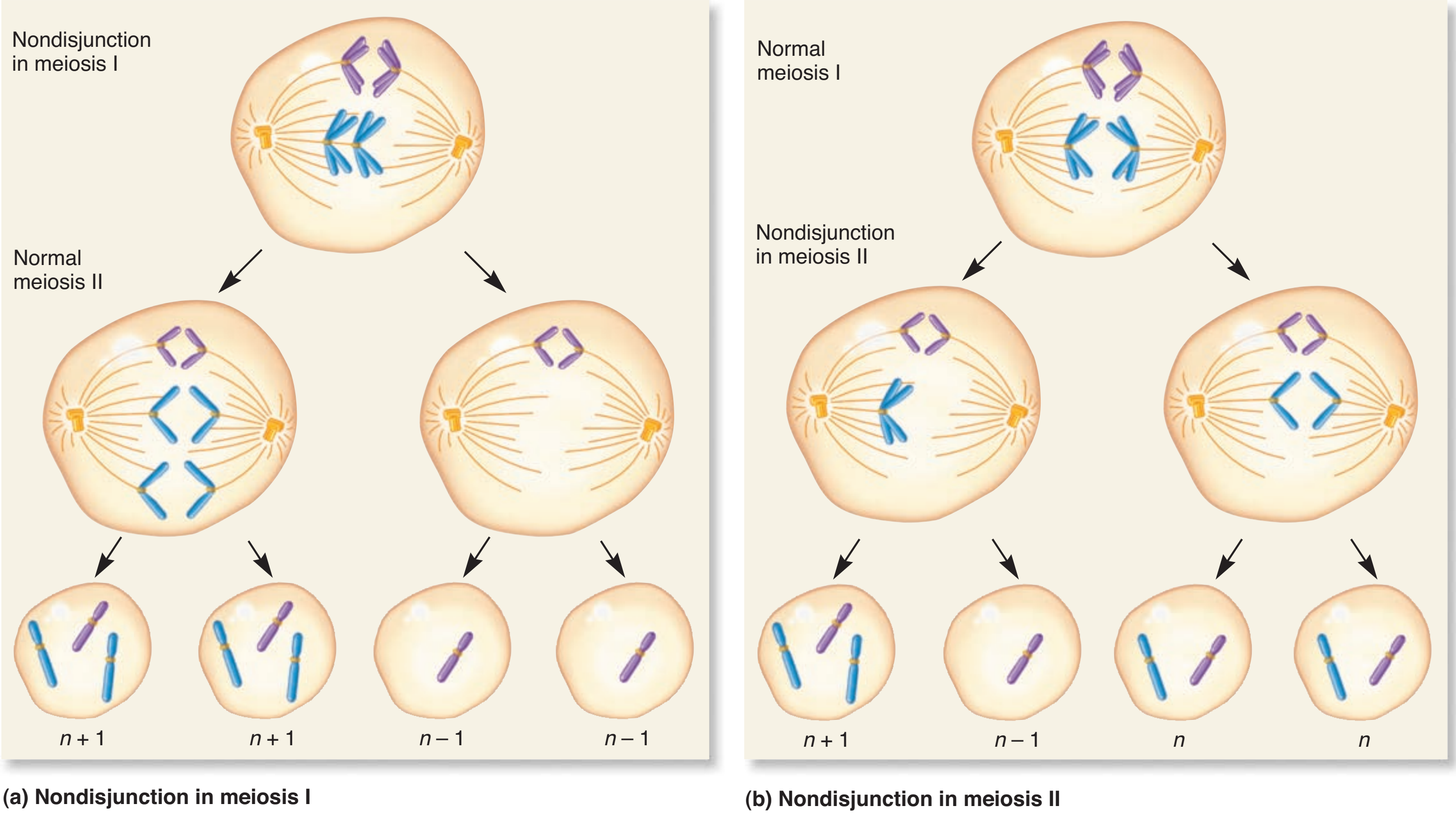
The normal process of meiosis is described in Chapter 2 (see Figures 2.12 and 2.13). Nondisjunction during meiosis can occur during anaphase of meiosis I or meiosis II. If it happens during meiosis I, an entire bivalent migrates to one pole (**Figure 8.22a**). Following the completion of meiosis, the four resulting haploid cells produced from this event are abnormal. If nondisjunction occurs during anaphase of meiosis II (**Figure 8.22b**), the net result is two abnormal and two normal haploid cells. If a gamete that is missing a chromosome is viable and participates in fertilization, the resulting offspring is monosomic for the missing chromosome. Alternatively, if a gamete carrying an extra chromosome unites with a normal gamete, the offspring will be trisomic.

In rare cases, all of the chromosomes can undergo nondisjunction and migrate to one of the daughter cells. The net result of **complete nondisjunction** is a diploid cell and a cell without any chromosomes. The cell without chromosomes is nonviable, but the diploid cell might participate in fertilization with a normal haploid gamete to produce a triploid individual. Therefore, complete nondisjunction can produce individuals that are polyploid.

Mitotic Nondisjunction or Chromosome Loss Can Produce a Patch of Tissue with an Altered Chromosome Number

Abnormalities in chromosome number occasionally occur after fertilization takes place. In this case, the abnormal event happens during mitosis rather than meiosis. One possibility is that the sister chromatids separate improperly, so one daughter cell has three copies of a chromosome, whereas the other daughter cell has only one (**Figure 8.23a**). Alternatively, the sister chromatids could separate during anaphase of mitosis, but one of the chromosomes could be improperly attached to the spindle, so it does not migrate to a pole (**Figure 8.23b**). A chromosome will be degraded if it is left outside the nucleus when the nuclear membrane re-forms. In this case, one of the daughter cells has two copies of that chromosome, whereas the other has only one.

When genetic abnormalities occur after fertilization, the organism contains a subset of cells that are genetically different from those of the rest of the organism. This condition is referred to as **mosaicism.** The size and location of the mosaic region depend on the timing and location of the original abnormal event. If a genetic alteration happens very early in the embryonic development of an organism, the abnormal cell will be the precursor for a large section of the organism. In the most extreme case, an abnormality can take place at the first mitotic division. As a bizarre example, consider a fertilized *Drosophila* egg that is XX. One of the X chromosomes may be lost during the first mitotic division, producing one daughter cell that is XX and one that is X0. Flies that are XX develop into females, and X0

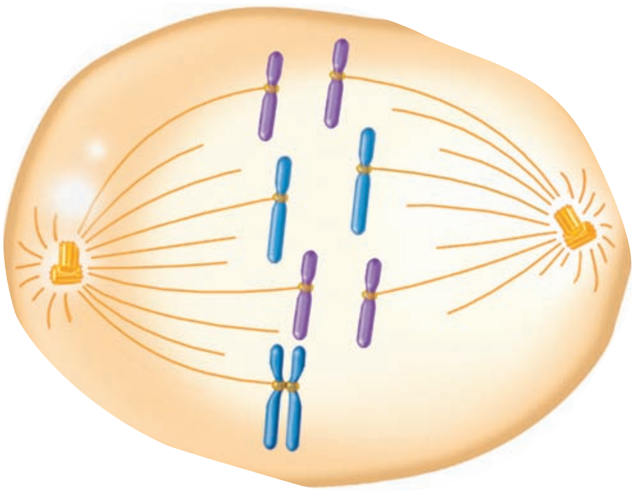


(a) Nondisjunction in meiosis I

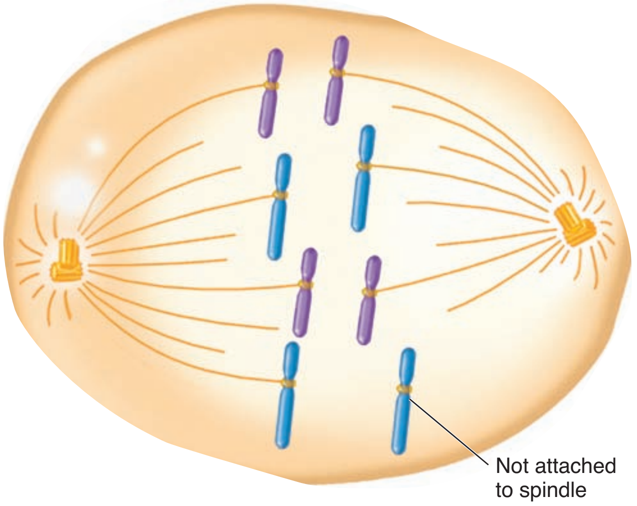
(b) Nondisjunction in meiosis II

FIGURE 8.22 **Nondisjunction during meiosis I and II.** The chromosomes shown in purple are behaving properly during meiosis I and II, so each cell receives one copy of this chromosome. The chromosomes shown in blue are not disjoining correctly. In **(a)**, nondisjunction occurred in meiosis I, so the resulting four cells receive either two copies of the blue chromosome or zero copies. In **(b)**, nondisjunction occurred during meiosis II, so one cell has two blue chromosomes and another cell has zero. The remaining two cells are normal.

Concept check: Explain what the word nondisjunction means.



(a) Mitotic nondisjunction



(b) Chromosome loss

FIGURE 8.23 **Nondisjunction and chromosome loss during mitosis in somatic cells.** **(a)** Mitotic nondisjunction produces a trisomic and a monosomic daughter cell. **(b)** Chromosome loss produces a normal and a monosomic daughter cell.

flies develop into males. Therefore, in this example, one-half of the organism becomes female and one-half becomes male! This peculiar and rare individual is referred to as a **bilateral gynandromorph** (**Figure 8.24**).

Changes in Euploidy Can Occur by Autopolyploidy, Alloploidy, and Allopolyploidy

Different mechanisms account for changes in the number of chromosome sets among natural populations of plants and animals (**Figure 8.25**). As previously mentioned, complete nondisjunction, due to a general defect in the spindle apparatus, can produce an individual with one or more extra sets of chromosomes. This individual is known as an **autopolyploid** (Figure 8.25a). The prefix *auto-* (meaning self) and the term *polyploid* (meaning many sets of chromosomes) refer to an increase in the number of chromosome sets within a single species.

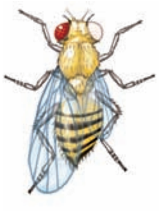


FIGURE 8.24 **A bilateral gynandromorph of *Drosophila melanogaster.***

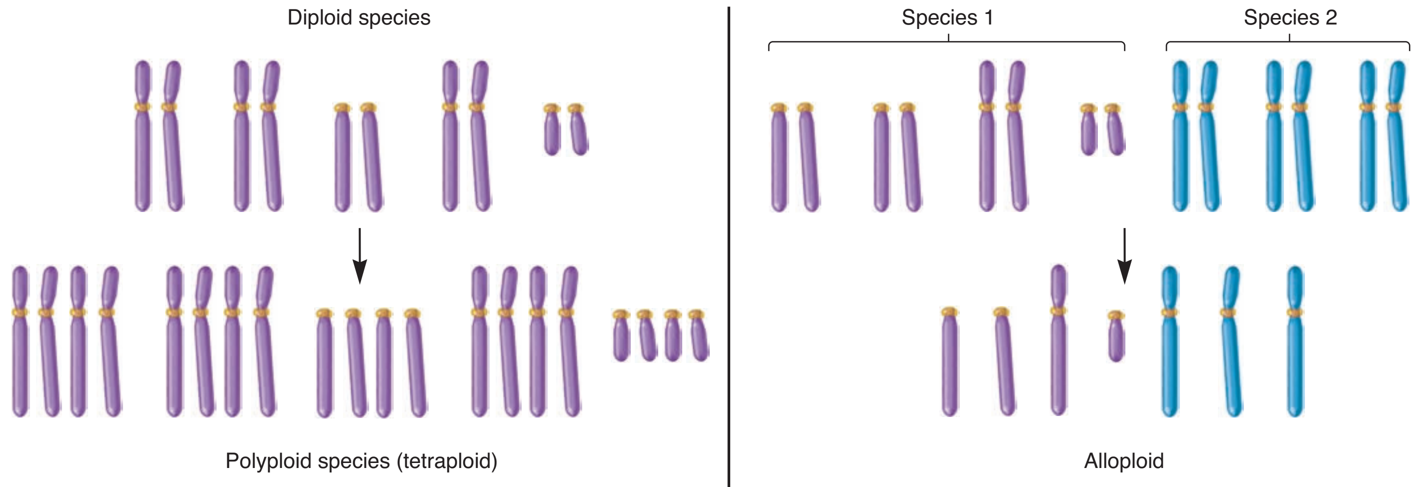
Genes → Traits In *Drosophila,* the ratio between the number of X chromosomes and sets of autosomes determines sex. This fly began as an XX female. One X chromosome carried the recessive white-eye and miniature wing alleles; the other X chromosome carried the wild-type alleles. The X chromosome carrying the wild-type alleles was lost from one of the cells during the first mitotic division, producing one XX cell and one X0 cell. The XX cell became the precursor for the left side of the fly, which developed as female. The X0 cell became the precursor for the other side of the fly, which developed as male with a white eye and a miniature wing.

Another common mechanism for change in chromosome number, called **alloploidy,** is a result of interspecies crosses (Figure 8.25b). An alloploid that has one set of chromosomes from two different species is called an **allodiploid.** This event is most likely to occur between species that are close evolutionary relatives. For example, closely related species of grasses may interbreed to produce allodiploids. As shown in Figure 8.25c, an **allopolyploid** contains two (or more) sets of chromosomes from two (or more) species. In this case, the **allotetraploid** contains two complete sets of chromosomes from two different species, for a total of four sets. Allotetraploids are also called amphidiploids. In nature, allotetraploids usually arise from allodiploids. This can occur when a somatic cell in an allodiploid undergoes complete nondisjunction to create an allotetraploid cell. In plants, such a cell can continue to grow and produce a section of the plant that is allotetraploid. If this part of the plant produces seeds by self-pollination, the seeds give rise to allotetraploid offspring. Cultivated wheat (refer back to Figure 8.20a) is a plant in which two species must have interbred to create an allotetraploid, and then a third species interbred with the allotetraploid to create an allohexaploid.

Allodiploids Are Often Sterile, but Allotetraploids Are More Likely to Be Fertile

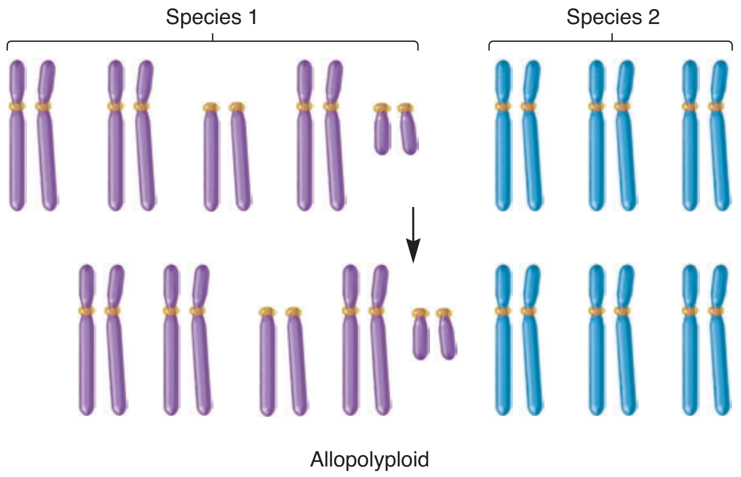
Geneticists are interested in the production of alloploids and allopolyploids as ways to generate interspecies hybrids with desirable traits. For example, if one species of grass can withstand hot temperatures and a closely related species is adapted to survive cold winters, a plant breeder may attempt to produce an interspecies hybrid that combines both qualities—good growth in the heat and survival through the winter. Such an alloploid may be desirable in climates with both hot summers and cold winters.

An important determinant of success in producing a fertile allodiploid is the degree of similarity of the different species' chromosomes. In two very closely related species, the number and types of chromosomes might be very similar. **Figure 8.26** shows a karyotype of an interspecies hybrid between the roan antelope (*Hippotragus equinus*) and the sable antelope (*Hippotragus niger*). As seen here, these two closely related species have the same number of chromosomes. The sizes and banding patterns of the chromosomes show that they correspond to one another.



(a) Autopolyploidy (tetraploid)

(b) Alloploidy (allodiploid)



(c) Allopolyploidy (allotetraploid)

FIGURE 8.25 **A comparison of autopolyploidy, alloploidy, and allopolyploidy.**

Concept check: *What is the key difference between autopolyploidy and allopolyploidy?*

For example, chromosome 1 from both species is fairly large, has very similar banding patterns, and carries many of the same genes. Evolutionarily related chromosomes from two different species are called **homeologous** chromosomes (not to be confused with homologous). This allodiploid is fertile because the homeologous chromosomes can properly synapse during meiosis to produce haploid gametes.

The critical relationship between chromosome pairing and fertility was first recognized by the Russian cytogeneticist Georgi Karpechenko in 1928. He crossed a radish (*Raphanus)* and a cabbage (*Brassica*), both of which are diploid and contain 18 chromosomes. Each of these organisms produces haploid cells containing 9 chromosomes. Therefore, the allodiploid produced from this interspecies cross contains 18 chromosomes. However, because the radish and cabbage are not closely related species, the nine *Raphanus* chromosomes are distinctly different from the nine *Brassica* chromosomes. During meiosis I, the radish and cabbage chromosomes cannot synapse with each other. This prevents the proper chromosome pairing and results in a high degree of aneuploidy (**Figure 8.27a**). Therefore, the radish–cabbage hybrid is sterile.

Among his strains of sterile alloploids, Karpechenko discovered that on rare occasions a plant produced a viable seed. When such seeds were planted and subjected to karyotyping, the plants were found to be allotetraploids with two sets of chromosomes from each of the two species. In the example shown in **Figure 8.27b**, the radish–cabbage allotetraploid contains 36 chromosomes instead of 18. The homologous chromosomes from each of the two species can synapse properly. When anaphase of meiosis I occurs, the pairs of synapsed chromosomes can disjoin equally to produce cells with 18 chromosomes each (a haploid set from the radish plus a haploid set from the cabbage). These cells can give rise to gametes with 18 chromosomes that combine with

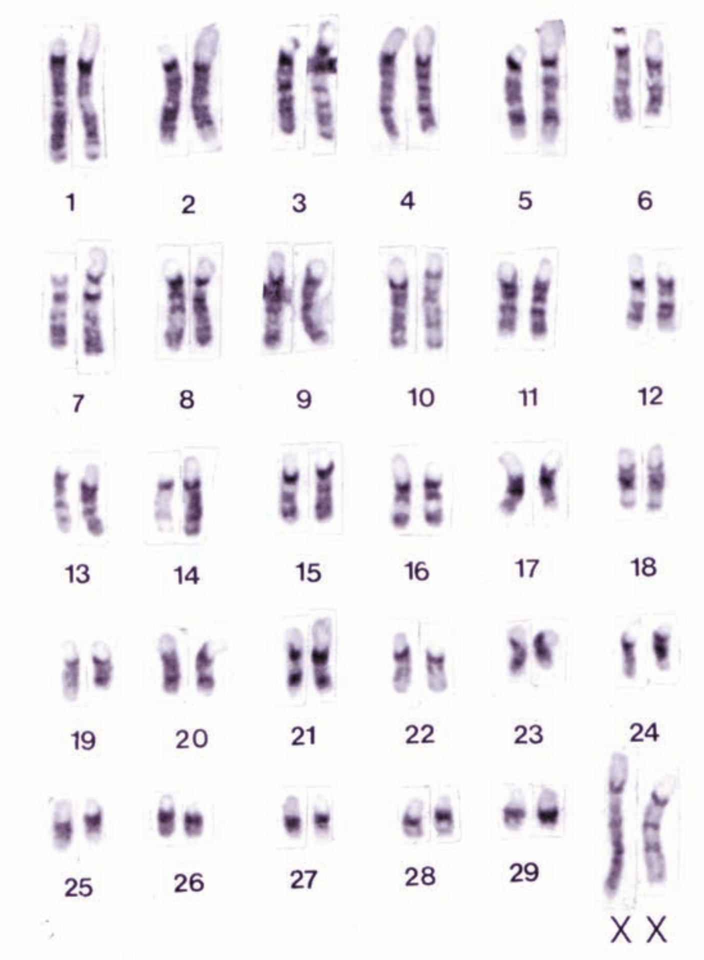
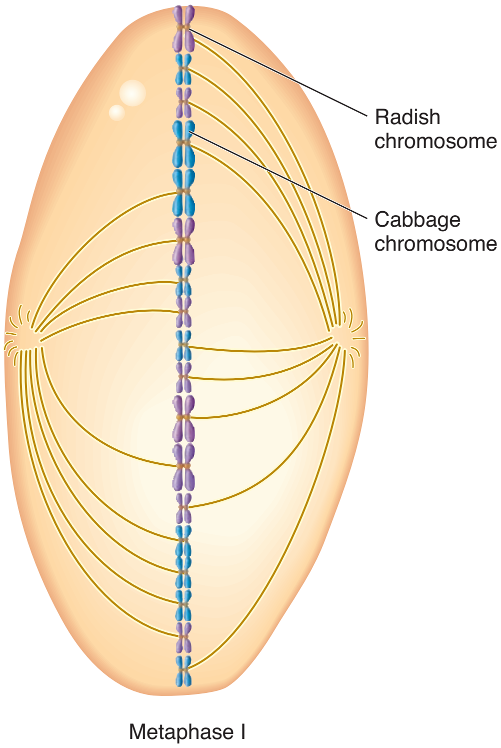
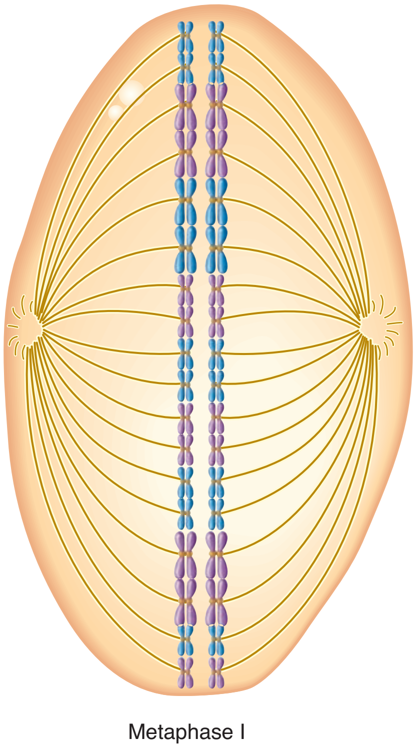


FIGURE 8.26 **The karyotype of a hybrid animal produced from two closely related antelope species.** In each chromosome pair in this karyotype, one chromosome was inherited from the roan antelope (*Hippotragus equinus*) and the other from the sable antelope (*Hippotragus niger*). When it is possible to distinguish the chromosomes from the two species, the roan chromosomes are shown on the left side of each pair.

each other to produce an allotetraploid containing 36 chromosomes. In this way, the allotetraploid is a fertile organism. Karpechenko's goal was to create a "vegetable for the masses" that would combine the nutritious roots of a radish with the flavorful leaves of a cabbage. Unfortunately, however, the allotetraploid had the leaves of the radish and the roots of the cabbage! Nevertheless, Karpechenko's scientific contribution was still important



(a) Allodiploid with a monoploid set from each species



(b) Allotetraploid with a diploid set from each species

FIGURE 8.27 **A comparison of metaphase I of meiosis in an allodiploid and an allotetraploid.** The purple chromosomes are from the radish (*Raphanus*), and the blue are from the cabbage (*Brassica*). **(a)** In the allodiploid, the radish and cabbage chromosomes have not synapsed and are randomly aligned during metaphase. **(b)** In the allotetraploid, the homologous radish and homologous cabbage chromosomes are properly aligned along the metaphase plate.

Concept check: Which of these two cells, an allodiploid and/or an allotetraploid, is/are likely to produce viable gametes? Explain why.

because he showed that it is possible to artificially produce a new self-perpetuating species of plant by creating an allotetraploid.

Modern plant breeders employ several different strategies to produce allotetraploids. One method is to start with two different tetraploid species. Because tetraploid plants make diploid gametes, the hybrid from this cross contains a diploid set from each species. Alternatively, if an allodiploid containing one set of chromosomes from each species already exists, a second approach is to create an allotetraploid by using agents that alter chromosome number. We will explore such experimental treatments next.

Experimental Treatments Can Promote Polyploidy

Because polyploid and allopolyploid plants often exhibit desirable traits, the development of polyploids is of considerable interest among plant breeders. Experimental studies on the ability of environmental agents to promote polyploidy began in the early 1900s. Since that time, various agents have been shown to promote nondisjunction, thereby leading to polyploidy. These include abrupt temperature changes during the initial stages of seedling growth and the treatment of plants with chemical agents that interfere with the formation of the spindle apparatus.

The drug colchicine is commonly used to promote polyploidy. Once inside the cell, colchicine binds to tubulin (a protein found in the spindle apparatus) and inhibits microtubule assembly, thereby interfering with normal chromosome segregation during mitosis or meiosis. In 1937, Alfred Blakeslee and Amos Avery applied colchicine to plant tissue and, at high doses, were able to cause complete mitotic nondisjunction and produce polyploidy in plant cells. Colchicine can be applied to seeds, young embryos, or rapidly growing regions of a plant (**Figure 8.28**). This application may produce aneuploidy, which is usually an undesirable outcome, but it often produces polyploid cells, which may grow faster than the surrounding diploid tissue. In a diploid plant, colchicine may cause complete mitotic nondisjunction, yielding tetraploid (4*n*) cells. As the tetraploid cells continue to divide, they generate a portion of the plant that is often morphologically distinguishable from the remainder. For example, a tetraploid stem may have a larger diameter and produce larger leaves and flowers.

Because individual plants can be propagated asexually from pieces of plant tissue (i.e., cuttings), the polyploid portion of the plant can be removed, treated with the proper growth hormones, and grown as a separate plant. Alternatively, the tetraploid region of a plant may have flowers that produce seeds by self-pollination. For many plant species, a tetraploid flower produces diploid pollen and eggs, which can combine to produce tetraploid offspring. In this way, the use of colchicine provides a straightforward method of producing polyploid strains of plants.

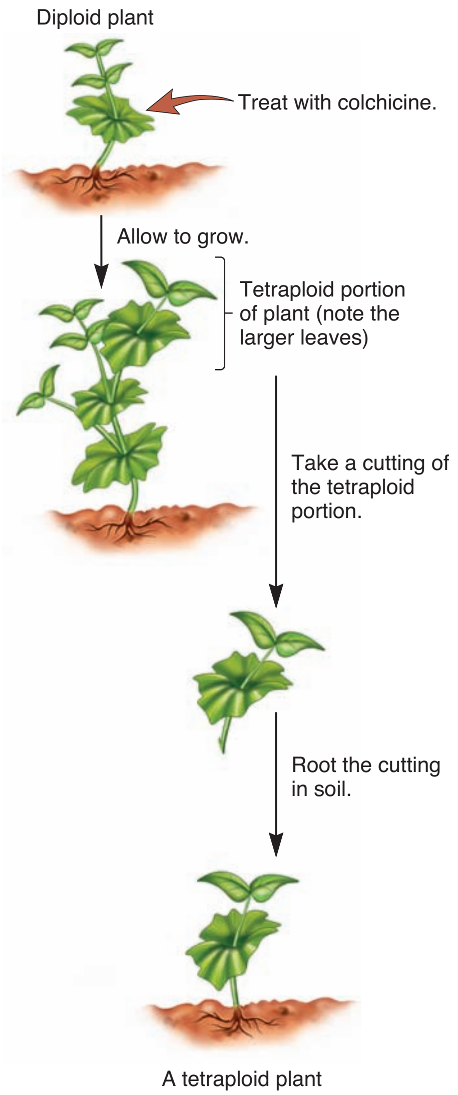

Diploid plant
Treat with colchicine.
Allow to grow.
Tetraploid portion of plant (note the larger leaves)
Take a cutting of the tetraploid portion.
Root the cutting in soil.
A tetraploid plant

FIGURE 8.28 Use of colchicine to promote polyploidy in plants. Colchicine interferes with the mitotic spindle apparatus and promotes nondisjunction. If complete nondisjunction occurs in a diploid cell, a daughter cell will be formed that is tetraploid. Such a tetraploid cell may continue to divide and produce a segment of the plant with more robust characteristics. This segment may be cut from the rest of the plant and rooted. In this way, a tetraploid plant can be propagated.

8.8 REVIEWING THE KEY CONCEPTS

- Meiotic nondisjunction, mitotic nondisjunction, and chromosome loss can result in changes in chromosome number (see Figures 8.22–8.24).
- Autopolyploidy is an increase in the number of sets within a single species. Interspecies matings result in alloploids. Alloploids with multiple sets of chromosomes from each species are allopolyploids (see Figures 8.25, 8.26).
- The chromosomes from allodiploids may not be able to pair up during meiosis, but the homologs in an allotetraploid can. For this reason, allotetraploids are usually fertile (see Figure 8.27).
- Various agents, including the drug colchicine, are used to promote nondisjunction, thereby producing polyploid plants (see Figure 8.28)

8.8 COMPREHENSION QUESTIONS

1. In a diploid species, complete nondisjunction during meiosis I may produce a viable cell that is
   a. trisomic.

b. haploid.
c. diploid.
d. triploid.

2. The somatic cells of an allotetraploid usually contain
   a. one set of chromosomes from four different species.
   b. two sets of chromosomes from two different species.
   c. four sets of chromosomes from one species.
   d. one set of chromosomes from two different species.

KEY TERMS

Page 155. genetic variation, allelic variation, cytogeneticist
Page 156. metacentric, submetacentric, acrocentric, telocentric, karyotype, G bands
Page 158. deletion, deficiency, duplication, inversion, translocation, simple translocation, reciprocal translocation
Page 159. terminal deletion, interstitial deletion, repetitive sequences, nonallelic homologous recombination
Page 160. gene duplication, gene family, homologous, paralogs
Page 161. copy number variation (CNV), segmental duplication
Page 162. pericentric inversion, paracentric inversion, position effect
Page 163. inversion heterozygote, inversion loop
Page 164. dicentric, dicentric bridge, acentric fragment, telomeres
Page 165. balanced translocations, unbalanced translocation, Robertsonian translocation, translocation cross
Page 167. semisterility
Page 168. euploid, triploid
Page 169. polyploid, tetraploid, aneuploidy, trisomic, monosomic
Page 170. nondisjunction
Page 171. haplodiploid, endopolyploidy
Page 172. polytene chromosome
Page 173. chromocenter
Page 174. meiotic nondisjunction, mitotic nondisjunction, alloploid
Page 175. complete nondisjunction, mosaicism
Page 176. bilateral gynandromorph, autopolyploid, alloploidy, allodiploid, allopolyploid, allotetraploid
Page 177. homeologous

CHAPTER SUMMARY

8.1 Microscopic Examination of Eukaryotic Chromosomes

- Among different species, natural variation exists with regard to chromosome structure and number. Three features of chromosomes that aid in their identification are centromere location, size, and banding patterns (see Figure 8.1).

8.2 Changes in Chromosome Structure: An Overview

- Variations in chromosome structure include deletions, duplications, inversions, and translocations (see Figure 8.2).

8.3 Deletions and Duplications

- Chromosome breaks can create terminal and interstitial deletions. Some deletions are associated with human genetic disorders such as cri-du-chat syndrome (see Figures 8.3, 8.4).
- Nonallelic homologous recombination can create gene duplications and deletions. Over time, gene duplications can lead to the formation of gene families, such as the globin gene family (see Figures 8.5–8.7).
- Copy number variation (CNV) is fairly common within a species. In humans, CNV is associated with certain diseases (see Figure 8.8).

8.4 Inversions and Translocations

- Inversions can be pericentric or paracentric. In an inversion heterozygote, crossing over within the inversion loop can create deletions and duplications in the resulting chromosomes (see Figures 8.9, 8.10).
- Two mechanisms that may produce translocations are chromosome breakage and subsequent repair, and nonhomologous crossing over (see Figure 8.11).
- A Robertsonian translocation results in a chromosome in which the long arms of two different acrocentric chromosomes become joined (see Figure 8.12).
- Familial Down syndrome is due to a translocation between chromosome 14 and 21 (see Figure 8.13).
- Individuals that carry a balanced translocation may have a high probability of producing unbalanced gametes, depending on the segregation pattern during meiosis I (see Figure 8.14).

8.5 Changes in Chromosome Number: An Overview

- Variation in the number of chromosomes may involve changes in the number of sets (euploidy) or changes in the number of particular chromosomes within a set (aneuploidy) (see Figure 8.15).

8.6 Variation in the Number of Chromosomes Within a Set: Aneuploidy

- Aneuploidy is often detrimental due to an imbalance in gene expression. Down syndrome, an example of an aneuploidy in humans, increases in frequency with maternal age (see Figures 8.16, 8.17, Table 8.1).

8.7 Variation in the Number of Sets of Chromosomes

- Among animals, variation in euploidy is relatively rare, though it does occur. Some tissues within an animal may exhibit polyploidy. An example is the polytene chromosomes found in salivary cells of *Drosophila* (see Figures 8.18, 8.19).

- Polyploidy in plants is relatively common and has found many advantages in agriculture. Triploid plants are usually seedless because they cannot segregate their chromosomes evenly during meiosis (see Figures 8.20, 8.21).

8.8 Mechanisms That Produce Variation in Chromosome Number

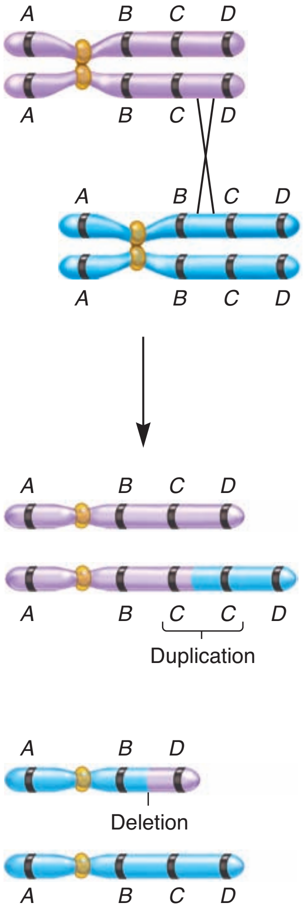
- Meiotic nondisjunction, mitotic nondisjunction, and chromosome loss can result in changes in chromosome number (see Figures 8.22–8.24).
- Autopolyploidy is an increase in the number of sets within a single species. Interspecies matings result in alloploids. Alloploids with multiple sets of chromosomes from each species are allopolyploids (see Figures 8.25, 8.26).
- The chromosomes from allodiploids may not be able to pair up during meiosis, but the homologs in an allotetraploid can. For this reason, allotetraploids are usually fertile (see Figure 8.27).
- Various agents, including the drug colchicine, are used to promote nondisjunction, thereby producing polyploid plants (see Figure 8.28).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe how a gene family is produced. Discuss the common and unique features of the family members in the globin gene family.

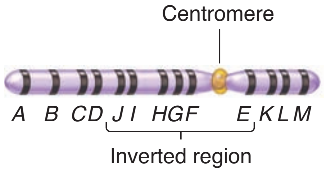
Answer: A gene family is produced when a single gene is copied one or more times by a gene duplication event. This duplication may occur by an abnormal (misaligned) crossover, which produces a chromosome with a deletion and another chromosome with a gene duplication.



Over time, this type of duplication may occur several times to produce many copies of a particular gene. In addition, translocations may move the duplicated genes to other chromosomes, so the members of the gene family may be dispersed among several different chromosomes. Eventually, each member of a gene family accumulates mutations, which may subtly alter their function.

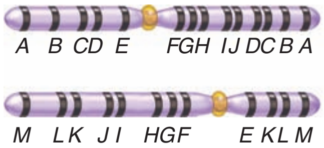
All members of the globin gene family bind oxygen. Myoglobin tends to bind it more tightly; therefore, it is good at storing oxygen. Hemoglobin binds it more loosely, so it can transport oxygen throughout the body (via red blood cells) and release it to the tissues that need oxygen. The polypeptides that form hemoglobins are predominantly expressed in red blood cells, whereas the myoglobin gene is expressed in many different cell types. The expression pattern of the globin genes changes during different stages of development. The ε- and ζ-globin genes are expressed in the early embryo. They are turned off near the end of the first trimester, and then the γ-globin genes exhibit their maximal expression during the second and third trimesters of gestation. Following birth, the γ-globin genes are silenced, and the β-globin gene is expressed for the rest of a person's life. These differences in the expression of the globin genes reflect the differences in the oxygen transport needs of humans during different stages of life. Overall, the evolution of gene families has resulted in gene products that are better suited to a particular tissue or stage of development. This has allowed a better "fine-tuning" of human traits.

S2. An inversion heterozygote has the following inverted chromosome:



What is the result if a crossover occurs between genes *F* and *G* on one inversion and one normal chromosome?

Answer: The resulting product is four chromosomes. One chromosome is normal, one is an inversion chromosome, and two chromosomes have duplications and deletions. The two duplicated or deficient chromosomes are shown here:



S3. In humans, the number of chromosomes per set equals 23. Even though the following conditions are lethal, what would be the total number of chromosomes for the following individuals?

A. Trisomy 22

B. Monosomy 11

C. Triploid individual

Answer:

A. 47 (the diploid number, 46, plus 1)

B. 45 (the diploid number, 46, minus 1)

C. 69 (3 times 23)

S4. A diploid species with 44 chromosomes (i.e., 22/set) is crossed to another diploid species with 38 chromosomes (i.e., 19/set). How many chromosomes are produced in an allodiploid or allotetraploid from this cross? Would you expect the offspring to be sterile or fertile?

Answer: An allodiploid would have 22 + 19 = 41 chromosomes. This individual would likely be sterile, because not all of the chromosomes would have homeologous partners to pair with during meiosis. This yields aneuploidy, which usually causes sterility. An allotetraploid would have 44 + 38 = 82 chromosomes. Because each chromosome would have a homologous partner, the allotetraploid would likely be fertile.

Conceptual Questions

C1. Which changes in chromosome structure cause a change in the total amount of genetic material, and which do not?

C2. Explain why small deletions and duplications are less likely to have a detrimental effect on an individual's phenotype than large ones. If a small deletion within a single chromosome happens to have a phenotypic effect, what would you conclude about the genes in this region?

C3. How does a chromosomal duplication occur?

C4. What is a gene family? How are gene families produced over time? With regard to gene function, what is the biological significance of a gene family?

C5. Following a gene duplication, two genes will accumulate different mutations, causing them to have slightly different sequences. In Figure 8.7, which pair of genes would you expect to have more similar sequences, α_1 and α_2 or ψ_{α_1} and α_2? Explain your answer.

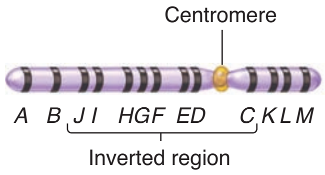
C6. Two chromosomes have the following order of genes:

Normal: *A B C* centromere *D E F G H I*

Abnormal: *A B G F E D* centromere *C H I*

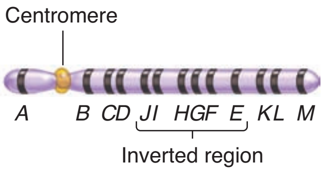
Does the abnormal chromosome have a pericentric or paracentric inversion? Draw a sketch showing how these two chromosomes would pair during prophase of meiosis I.

C7. An inversion heterozygote has the following inverted chromosome:



What would be the products if a crossover occurred between genes *H* and *I* on one inverted and one normal chromosome?

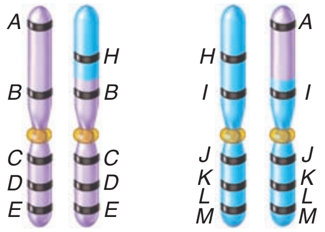
C8. An inversion heterozygote has the following inverted chromosome:



What would be the products if a crossover occurred between genes *H* and *I* on one inverted and one normal chromosome?

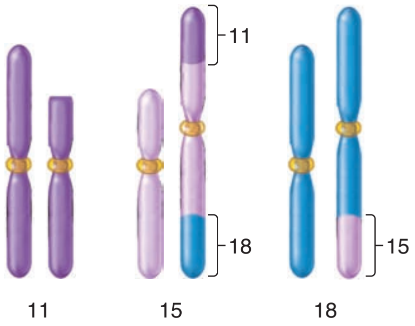
C9. Explain why inversions and reciprocal translocations do not usually cause a phenotypic effect. In a few cases, however, they do. Explain how.

C10. An individual has the following reciprocal translocation:



What would be the outcome of alternate and adjacent-1 segregation?

C11. A phenotypically normal individual has the following combinations of abnormal chromosomes:



The normal chromosomes are shown on the left of each pair. Suggest a series of events (breaks, translocations, crossovers, etc.) that may have produced this combination of chromosomes.

C12. Two phenotypically normal parents produce a phenotypically abnormal child in which chromosome 5 is missing part of its long arm but has a piece of chromosome 7 attached to it. The child also has one normal copy of chromosome 5 and two normal copies of chromosome 7. With regard to chromosomes 5 and 7, what do you think are the chromosomal compositions of the parents?

C13. With regard to the segregation of centromeres, why is adjacent-2 segregation less frequent than alternate or adjacent-1 segregation?

C14. Which of the following types of chromosomal changes would you expect to have phenotypic consequences? Explain your choices.

A. Pericentric inversion

B. Reciprocal translocation

C. Deletion

D. Unbalanced translocation

C15. Explain why a translocation cross occurs during metaphase of meiosis I when a cell contains a reciprocal translocation.

C16. A phenotypically abnormal individual has a phenotypically normal father with an inversion on one copy of chromosome 7 and a normal mother without any changes in chromosome structure. The order of genes along chromosome 7 in the father is as follows:

R T D M centromere *P U X Z C* (normal chromosome 7)

R T D U P centromere *M X Z C* (inverted chromosome 7)

The phenotypically abnormal offspring has a chromosome 7 with the following order of genes:

R T D M centromere *P U D T R*

With a sketch, explain how this chromosome was formed. In your answer, explain where the crossover occurred (i.e., between which two genes).

C17. A diploid fruit fly has eight chromosomes. How many total chromosomes would be found in the following flies?

A. Tetraploid

B. Trisomy 2

C. Monosomy 3

D. $3n$

E. $4n + 1$

C18. A person is born with one X chromosome, zero Y chromosomes, trisomy 21, and two copies of the other chromosomes. How many chromosomes does this person have altogether? Explain whether this person is euploid or aneuploid.

C19. Two phenotypically unaffected parents produce two children with familial Down syndrome. With regard to chromosomes 14 and 21, what are the chromosomal compositions of the parents?

C20. Aneuploidy is typically detrimental, whereas polyploidy is sometimes beneficial, particularly in plants. Discuss why you think this is the case.

C21. Explain how aneuploidy, deletions, and duplications cause genetic imbalances. Why do you think that deletions and monosomies are more detrimental than duplications and trisomies?

C22. Why do you think that human trisomies 13, 18, and 21 can survive but the other trisomies are lethal? Even though X chromosomes are large, aneuploidies of this chromosome are also tolerated. Explain why.

C23. A zookeeper has collected a male and female lizard that look like they belong to the same species. They mate with each other and produce phenotypically normal offspring. However, the offspring are sterile. Suggest one or more explanations for their sterility.

C24. What is endopolyploidy? What is its biological significance?

C25. What is mosaicism? How is it produced?

C26. Explain how polytene chromosomes of *Drosophila melanogaster* are produced and how they form a six-armed structure.

C27. Describe some of the advantages of polyploid plants. What are the consequences of having an odd number of chromosome sets?

C28. A diploid fruit fly has eight chromosomes. Which of the following terms should not be used to describe a fruit fly with four sets of chromosomes?

A. Polyploid

B. Aneuploid

C. Euploid

D. Tetraploid

E. $4n$

C29. Which of the following terms should not be used to describe a human with three copies of chromosome 12?

A. Polyploid

B. Triploid

C. Aneuploid

D. Euploid

E. $2n + 1$

F. Trisomy 12

C30. The kidney bean, *Phaseolus vulgaris,* is a diploid species containing a total of 22 chromosomes in somatic cells. How many possible types of trisomic individuals could be produced in this species?

C31. The karyotype of a young girl who is affected with familial Down syndrome revealed a total of 46 chromosomes. Her older brother, however, who is phenotypically unaffected, actually had 45 chromosomes. Explain how this could happen. What would you expect to be the chromosomal number in the parents of these two children?

C32. A triploid plant has 18 chromosomes (i.e., 6 chromosomes per set). If we assume a gamete has an equal probability of receiving one or two copies of each of the six types of chromosome, what are the odds of this plant producing a monoploid or a diploid gamete? What are the odds of producing an aneuploid gamete? If the plant is allowed to self-fertilize, what are the odds of producing a euploid offspring?

C33. Describe three naturally occurring ways that the chromosome number can change.

C34. Meiotic nondisjunction is much more likely than mitotic nondisjunction. Based on this observation, would you conclude that meiotic nondisjunction is usually due to nondisjunction during meiosis I or meiosis II? Explain your reasoning.

C35. A woman who is heterozygous, *Bb*, has brown eyes. *B* (brown) is a dominant allele, and *b* (blue) is recessive. In one of her eyes, however, there is a patch of blue color. Give three different explanations for how this might have occurred.

C36. What is an allodiploid? What factor determines the fertility of an allodiploid? Why are allotetraploids more likely to be fertile?

C37. What are homeologous chromosomes?

C38. Meiotic nondisjunction usually occurs during meiosis I. What is not separating properly: bivalents or sister chromatids? What is not separating properly during mitotic nondisjunction?

C39. Table 8.1 shows that Turner syndrome occurs when an individual inherits one X chromosome but lacks a second sex chromosome. Can Turner syndrome be due to nondisjunction during oogenesis, spermatogenesis, or both? If a phenotypically normal couple has a color-blind child (due to a recessive X-linked allele) with Turner syndrome, did nondisjunction occur during oogenesis or spermatogenesis in this child's parents? Explain your answer.

C40. Male honeybees, which are monoploid, produce sperm by meiosis. Explain what unusual event (compared with other animals) must occur during spermatogenesis in honeybees to produce sperm. Does this unusual event occur during meiosis I or meiosis II?

Application and Experimental Questions

E1. With regard to the analysis of chromosome structure, explain the experimental advantage that polytene chromosomes offer. Discuss why changes in chromosome structure are more easily detected in polytene chromosomes than in ordinary (nonpolytene) chromosomes.

E2. Describe how colchicine can be used to alter chromosome number.

E3. Female fruit flies homozygous for the X-linked white-eye allele are crossed to males with red eyes. On very rare occasions, an offspring is a male with red eyes. Assuming these rare offspring are not due to a new mutation in one of the mother's X chromosomes that converted the white-eye allele into a red-eye allele, explain how this red-eyed male arose.

E4. A cytogeneticist has collected tissue samples from members of the same butterfly species. Some of the butterflies were located in Canada, and others were found in Mexico. Upon karyotyping, the cytogeneticist discovered that chromosome 5 of the Canadian butterflies had a large inversion compared with the Mexican butterflies. The Canadian butterflies were inversion homozygotes, whereas the Mexican butterflies had two normal copies of chromosome 5.

A. Explain whether a mating between the Canadian and Mexican butterflies would produce phenotypically normal offspring.

B. Explain whether the offspring of a cross between Canadian and Mexican butterflies would be fertile.

E5. While conducting field studies on a chain of islands, you decide to karyotype two phenotypically identical groups of turtles, which are found on different islands. The turtles on one island have 24 chromosomes, but those on another island have 48 chromosomes. How would you explain this observation? How do you think the turtles with 48 chromosomes came into being? If you mated the two types of turtles, would you expect their offspring to be phenotypically normal? Would you expect them to be fertile? Explain.

E6. It is an exciting time to be a plant breeder because so many options are available for the development of new types of agriculturally useful plants. Let's suppose you wish to develop a seedless tomato that could grow in a very hot climate and is resistant to a viral pathogen that commonly infects tomato plants. At your disposal, you have a seed-bearing tomato strain that is heat-resistant and produces great-tasting tomatoes. You also have a wild strain of tomato plants (which have lousy-tasting tomatoes) that is resistant to the viral pathogen. Suggest a series of steps you might follow to produce a great-tasting, seedless tomato that is resistant to heat and the viral pathogen.

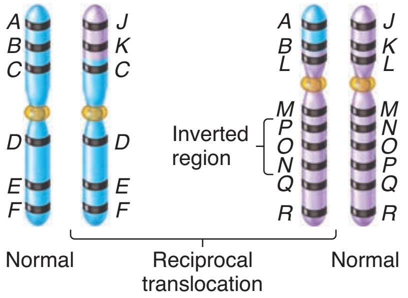
E7. What is a G-band? Discuss how G-bands are useful in the analysis of chromosome structure.

E8. A female fruit fly contains one normal X chromosome and one X chromosome with a deletion. The deletion is in the middle of the X chromosome and is about 10% of the entire length of the X chromosome. If you stained and observed the chromosomes of this female fly in salivary gland cells, draw a picture of the polytene arm of the X chromosome. Explain your drawing.

E9. Describe two different experimental strategies to create an allotetraploid from two different diploid species of plants.

Questions for Student Discussion/Collaboration

1. A chromosome involved in a reciprocal translocation also has an inversion. In addition, the cell contains two normal chromosomes.



Make a drawing that shows how these chromosomes will pair during metaphase of meiosis I.

2. Besides the ones mentioned in this textbook, look for other examples of variations in euploidy. Perhaps you might look in more advanced textbooks concerning population genetics, ecology, or the like. Discuss the phenotypic consequences of these changes.

3. Cell biology textbooks often discuss cellular proteins encoded by genes that are members of a gene family. Examples of such proteins include myosins and glucose transporters. Look through a cell biology textbook and identify some proteins encoded by members of gene families. Discuss the importance of gene families at the cellular level.

4. Discuss how variation in chromosome number has been useful in agriculture.

Answers to Comprehension Questions

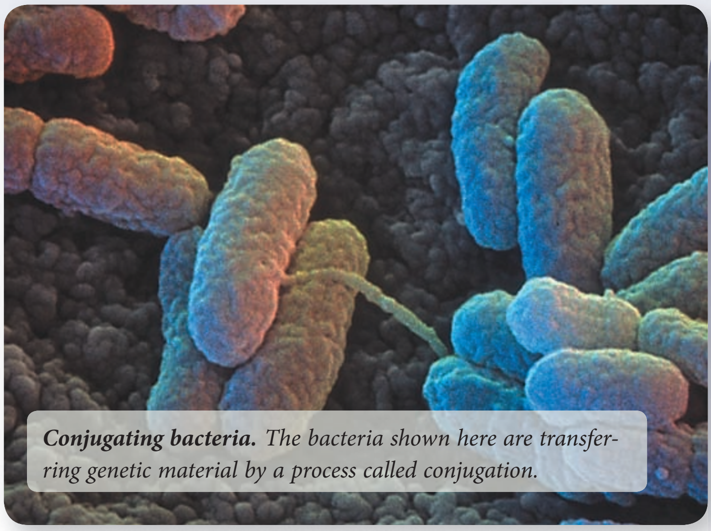
8.1: c, d
8.2: c
8.3: a, d
8.4: b, d, d
8.5: b
8.6: a, d
8.7: c, b
8.8: c, b

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

9



Conjugating bacteria. *The bacteria shown here are transferring genetic material by a process called conjugation.*

CHAPTER OUTLINE

9.1 Overview of Genetic Transfer in Bacteria

9.2 Bacterial Conjugation

9.3 Conjugation and Mapping via *Hfr* Strains

9.4 Bacterial Transduction

9.5 Bacterial Transformation

GENETICS OF BACTERIA

Thus far, in Chapters 1 through 8, we primarily focused on eukaryotes, which include protists, fungi, plants, and animals. In Chapters 9 and 10, we will turn our attention to bacteria and viruses, respectively. One reason why researchers are so interested in bacteria and viruses is related to their impact on health. Infectious diseases caused by these agents are a leading cause of human death, accounting for one-quarter to one-third of deaths worldwide. The spread of infectious diseases results from human behavior, and in recent times has been accelerated by increased trade and travel and the inappropriate use of antibiotic drugs. Although the incidence of fatal infectious diseases in the United States is relatively low compared with the worldwide average, an alarming increase in more deadly strains of bacteria and viruses has occurred over the past few decades. Since 1980, the number of deaths in the United States due to infectious diseases has approximately doubled.

In this chapter, we will focus on the genetic analysis of bacteria. Like their eukaryotic counterparts, bacteria often possess allelic differences that affect their cellular traits. Common allelic variations among bacteria that are readily discernible involve traits such as sensitivity to antibiotics and differences in their nutrient requirements for growth. In these cases, the allelic differences are between different strains of bacteria, because any given bacterium is usually haploid for a particular gene. In fact, the haploid nature of bacteria is a feature that makes it easier to identify mutations that produce phenotypes such as altered nutritional requirements. Loss-of-function mutations, which are often recessive in diploid eukaryotes, are not masked by dominant alleles in haploid species. Throughout this chapter, we will consider interesting experiments that examine bacterial strains with allelic differences.

Compared with eukaryotes, another striking difference in prokaryotic species is their mode of reproduction. Because bacteria reproduce asexually, researchers do not use crosses in the genetic analysis of bacterial species. Instead, they rely on a similar mechanism, called genetic transfer, in which a segment of bacterial DNA is transferred from one bacterium to another. In this chapter, we will explore different routes of genetic transfer, and see how researchers have used genetic transfer to map the locations of genes along the chromosome of certain bacterial species.

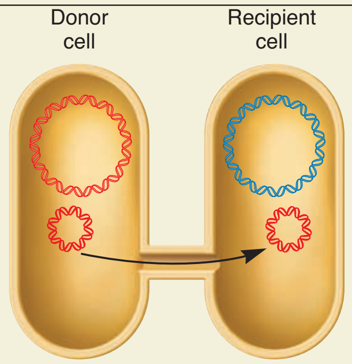
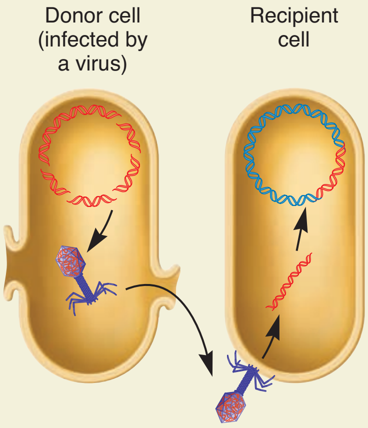
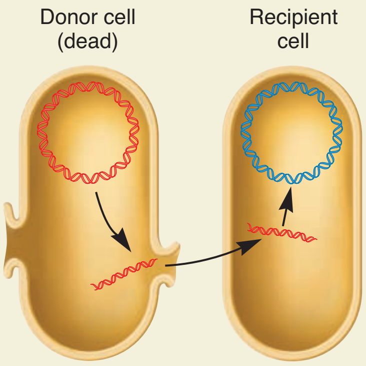
9.1 OVERVIEW OF GENETIC TRANSFER IN BACTERIA

In Chapter 2, we considered how bacteria reproduce asexually by a process called binary fission (refer back to Figure 2.4). This process produces two daughter cells that are genetically identical to the mother cell. In addition, a segment of DNA from one bacterial cell can occasionally be transferred to another cell. This process is called **genetic transfer.** Why is genetic transfer an advantage? Like sexual reproduction in eukaryotes, genetic transfer in bacteria is thought to enhance the genetic diversity of bacterial species. For example, a bacterial cell carrying a gene that provides antibiotic resistance may transfer this gene to another bacterial cell, allowing that bacterial cell to survive exposure to the antibiotic.

Bacteria can naturally transfer genetic material in three ways (Table 9.1). The first route, known as **conjugation,** involves a direct physical interaction between two bacterial cells. One bacterium acts as a donor and transfers genetic material to a recipient cell. A second means of transfer is called **transduction.** This occurs when a virus infects a bacterium and then transfers bacterial genetic material from that bacterium to another. The third mode of genetic transfer is **transformation.** In this case, genetic material is released into the environment when a bacterial cell dies. This material then binds to a living bacterial cell, which can take it up. These three mechanisms of genetic transfer have been extensively investigated

TABLE **9.1**

Three Mechanisms of Genetic Transfer Found in Bacteria

Mechanism		**Description**
Conjugation		Requires direct contact between a donor and a recipient cell. The donor cell transfers a strand of DNA to the recipient. In the example shown here, DNA known as a plasmid is transferred to the recipient cell.
Transduction		When a virus infects a donor cell, it incorporates a fragment of bacterial chromosomal DNA into a newly made virus particle. The virus then transfers this fragment of DNA to a recipient cell, which incorporates the DNA into its chromosome by recombination.
Transformation		When a bacterial cell dies, it releases a fragment of its DNA into the environment. This DNA fragment is taken up by a recipient cell, which incorporates the DNA into its chromosome by recombination.

in research laboratories, and their molecular mechanisms continue to be studied with great interest. In later sections, we will examine these three systems of genetic transfer in greater detail.

We will also learn how genetic transfer between bacterial cells has provided unique ways to accurately map bacterial genes. The mapping methods described in this chapter have been largely replaced by molecular approaches described in Chapter 21. Even so, the mapping of bacterial genes serves to illuminate the mechanisms by which genes are transferred between bacterial cells and also helps us to appreciate the strategies of newer mapping approaches.

9.1 REVIEWING THE KEY CONCEPTS

- Three general mechanisms for genetic transfer in various species of bacteria are conjugation, transduction, and transformation (see Table 9.1).

9.1 COMPREHENSION QUESTIONS

1. A form of genetic transfer that involves the uptake of a fragment of DNA from the environment is called
   a. conjugation.
   b. transduction.
   c. transformation.
   d. all of the above.

9.2 BACTERIAL CONJUGATION

As described briefly in Table 9.1, conjugation involves the direct transfer of genetic material from one bacterial cell to another. In this section, we will examine the steps in this process at the molecular and cellular levels.

Bacteria Can Transfer Genetic Material During Conjugation

The natural ability of some bacteria to transfer genetic material between each other was first recognized by Joshua Lederberg and Edward Tatum in 1946. They were studying strains of *Escherichia coli* that had different nutritional requirements for growth. A **minimal medium** is a growth medium that contains the essential nutrients for a wild-type (nonmutant) bacterial species to grow. Researchers often study bacterial strains that harbor mutations and cannot grow on minimal media. A strain that cannot synthesize a particular nutrient and needs that nutrient to be supplemented in its growth medium is called an **auxotroph.** For example, a strain that cannot make the amino acid methionine does not grow on a minimal medium. Such a strain needs to have methionine added to its growth medium and is called a methionine auxotroph. By comparison, a strain that can make this amino acid is termed a methionine prototroph. A **prototroph** does not need this nutrient in its growth medium.

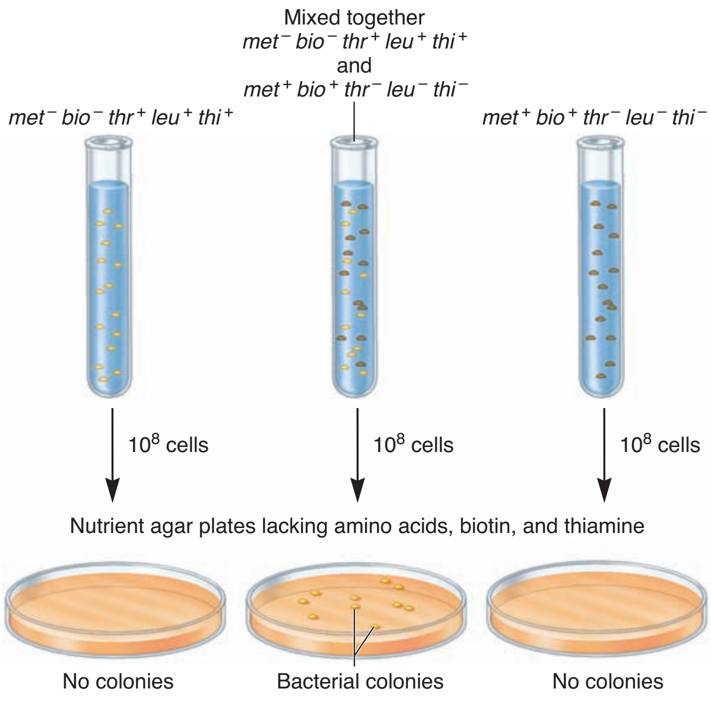

Mixed together
$met^- \; bio^- \; thr^+ \; leu^+ \; thi^+$
and
$met^+ \; bio^+ \; thr^- \; leu^- \; thi^-$
$met^- \; bio^- \; thr^+ \; leu^+ \; thi^+$
$met^+ \; bio^+ \; thr^- \; leu^- \; thi^-$
10^8 cells
10^8 cells
10^8 cells
Nutrient agar plates lacking amino acids, biotin, and thiamine
No colonies
Bacterial colonies
No colonies

FIGURE 9.1 Experiment of Lederberg and Tatum demonstrating genetic transfer during conjugation in *E. coli*. When plated on a growth medium lacking amino acids, biotin, and thiamine, the $met^- \; bio^- \; thr^+ \; leu^+ \; thi^+$ or $met^+ \; bio^+ \; thr^- \; leu^- \; thi^-$ strains were unable to grow. However, if they were mixed together and then plated, some colonies were observed. These colonies were due to the transfer of genetic material between these two strains by conjugation. *Note:* In bacteria, it is common to give genes a three-letter name (shown in italics) that is related to the function of the gene. A plus superscript ($^+$) indicates a functional gene, and a minus superscript ($^-$) indicates a mutation that has caused the gene or gene product to be inactive. In some cases, several genes have related functions. These may have the same three-letter name followed by different capital letters. For example, different genes involved with leucine biosynthesis may be called *leuA*, *leuB*, *leuC*, and so on. In the experiment described in Figure 9.1, the genes involved in leucine biosynthesis had not been distinguished, so the gene is simply referred to as leu^+ (for a functional gene) and leu^- (for a nonfunctional gene).

Concept check: Describe how gene transfer can explain the growth of colonies on the middle plate.

The experiment in **Figure 9.1** considers one *E. coli* strain, designated $met^- \; bio^- \; thr^+ \; leu^+ \; thi^+$, which required one amino acid, methionine (met), and one vitamin, biotin (bio), in order to grow. This strain did not require the amino acids threonine (thr) or leucine (leu), or the vitamin thiamine (thi) for growth. Another *E. coli* strain, designated $met^+ \; bio^+ \; thr^- \; leu^- \; thi^-$, had just the opposite requirements. It was an auxotroph for threonine, leucine, and thiamine, but a prototroph for methionine and biotin. These differences in nutritional requirements correspond to variations in the genetic material of the two strains. The first strain had two defective genes encoding enzymes necessary for methionine and biotin synthesis. The second strain contained three defective genes required to make threonine, leucine, and thiamine.

Figure 9.1 compares the results when the two strains were mixed together and when they were not mixed. Without mixing, about 100 million (10^8) $met^- \; bio^- \; thr^+ \; leu^+ \; thi^+$ cells were applied to plates on a growth medium lacking amino acids, biotin, and thiamine; no colonies were observed to grow. This result is expected because the medium did not contain methionine or biotin. Likewise, when 10^8 $met^+ \; bio^+ \; thr^- \; leu^- \; thi^-$ cells were plated, no colonies were observed because threonine, leucine, and thiamine were missing from this growth medium. However, when the two strains were mixed together and then 10^8 cells plated, approximately 10 bacterial colonies formed. A bacterial colony is derived from a single bacterial cell by many successive cell divisions. Because bacterial colonies were observed to grow, the genotype of the cells within these colonies must have been $met^+\ bio^+\ thr^+\ leu^+\ thi^+$. How could this genotype occur? Because no colonies were observed on either plate in which the two strains were not mixed, Lederberg and Tatum concluded that it was not due to mutations that converted $met^-\ bio^-$ to $met^+\ bio^+$ or to mutations that converted $thr^-\ leu^-\ thi^-$ to $thr^+\ leu^+\ thi^+$. Instead, they hypothesized that some genetic material was transferred between the two strains. One possibility is that the genetic material providing the ability to synthesize methionine and biotin ($met^+\ bio^+$) was transferred to the $met^-\ bio^-\ thr^+\ leu^+\ thi^+$ strain. Alternatively, the ability to synthesize threonine, leucine, and thiamine ($thr^+\ leu^+\ thi^+$) may have been transferred to the $met^+\ bio^+\ thr^-\ leu^-\ thi^-$ cells. The results of this experiment did not distinguish between these two possibilities.

Conjugation Requires Direct Physical Contact

In 1950, Bernard Davis conducted experiments showing that two strains of bacteria must make physical contact with each other to transfer genetic material. The apparatus he used, known as a U-tube, is shown in **Figure 9.2**. At the bottom of the U-tube is a filter with pores small enough to allow the passage of genetic material (i.e., DNA molecules) but too small to permit the passage of bacterial cells. On one side of the filter, Davis added a bacterial strain with a certain combination of nutritional requirements (the $met^-\ bio^-\ thr^+\ leu^+\ thi^+$ strain). On the other side, he added a different bacterial strain (the $met^+\ bio^+\ thr^-\ leu^-\ thi^-$ strain). The application of alternating pressure and suction promoted the movement of liquid through the filter. Because the bacteria were too large to pass through the pores, the movement of liquid did not allow the two types of bacterial strains to mix with each other. However, any genetic material that was released from a bacterium could pass through the filter.

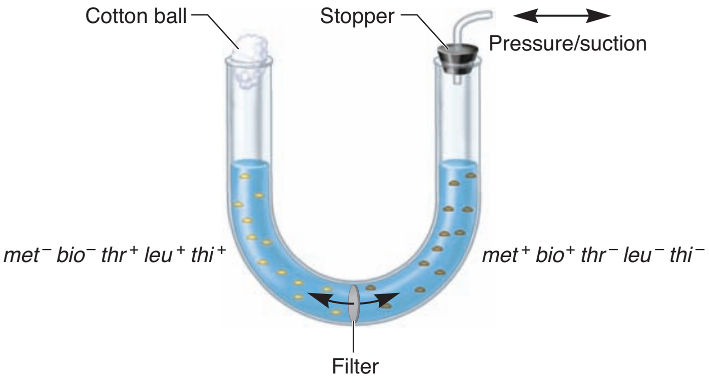

Cotton ball
Stopper
Pressure/suction
$met^-\ bio^-\ thr^+\ leu^+\ thi^+$
$met^+\ bio^+\ thr^-\ leu^-\ thi^-$
Filter

FIGURE 9.2 **A U-tube apparatus like that used by Davis.** The fluid in the tube is forced through the filter by alternating suction and pressure. However, the pores in the filter are too small for the passage of bacteria.

Concept check: With regard to the mechanism of conjugation, what is the purpose of using a U-tube?

After incubation in a U-tube, bacteria from either side of the tube were placed on a medium that could select for the growth of cells that were $met^+\ bio^+\ thr^+\ leu^+\ thi^+$. This selective medium lacked methionine, biotin, threonine, leucine, and thiamine, but contained all other nutrients essential for growth. In this case, no bacterial colonies grew on the plates. The experiment showed that, without physical contact, the two bacterial strains did not transfer genetic material to one another. The term *conjugation* is now used to describe the natural process of genetic transfer between bacterial cells that requires direct cell-to-cell contact. Many, but not all, species of bacteria can conjugate.

An F^+ Strain Transfers an F factor to an F^- Strain During Conjugation

Working independently in the early 1950s, Joshua and Esther Lederberg, William Hayes, and Luca Cavalli-Sforza discovered that only certain bacterial strains can act as donors of genetic material. For example, only about 5% of natural isolates of *E. coli* can act as donor strains. Research studies showed that a strain incapable of acting as a donor could subsequently be converted to a donor strain after being mixed with another donor strain. Hayes correctly proposed that donor strains contain a fertility factor that can be transferred to conjugation-defective strains to make them conjugation-proficient.

We now know that certain donor strains of *E. coli* contain a small circular segment of genetic material known as an **F factor** (for fertility factor) in addition to their circular chromosome. Strains of *E. coli* that contain an F factor are designated F^+, whereas strains without F factors are termed F^-. In recent years, the molecular details of the conjugation process have been extensively studied. Though the mechanisms vary somewhat from one bacterial species to another, some general themes have emerged. F factors carry several genes that are required for conjugation to occur. For example, **Figure 9.3** shows the arrangement of genes on the F factor found in certain strains of *E. coli*. The functions of the proteins encoded by these genes are needed to transfer a strand of DNA from the donor cell to a recipient cell.

Figure 9.4a describes the molecular events that occur during conjugation in *E. coli*. Contact between donor and recipient cells is a key step that initiates the conjugation process. **Sex pili** (singular: **pilus**) are made by F^+ strains (**Figure 9.4b**). The gene encoding the pilin protein (*traA*) is located on the F factor. The pili act as attachment sites that promote the binding of bacteria to each other. In this way, an F^+ strain makes physical contact with an F^- strain. In certain species, such as *E. coli*, long pili project from F^+ cells and attempt to make contact with nearby F^- cells. Once contact is made, the pili shorten and thereby draw the donor and recipient cells closer together. A **conjugation bridge** is then formed between the two cells, which provides a passageway for DNA transfer.

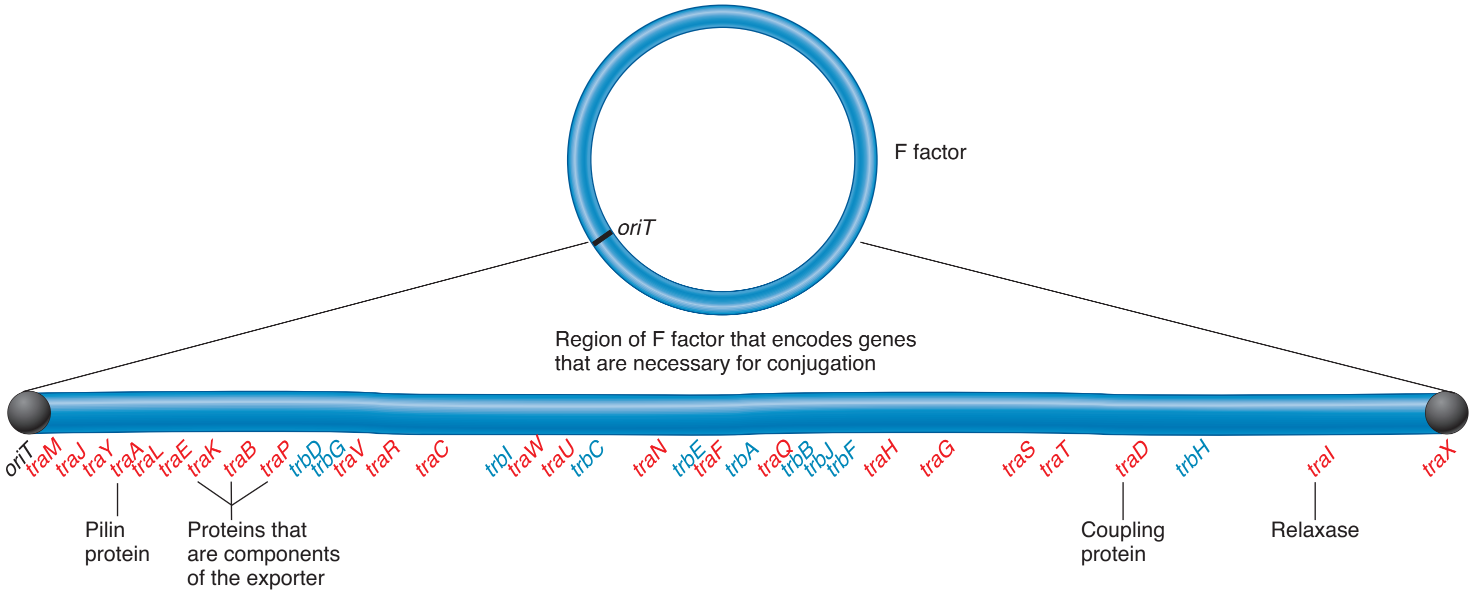


FIGURE 9.3 **Genes on the F factor that play a role during conjugation.** A region of the F factor contains a few dozen genes that play a role in the conjugative process. Because they play a role in the transfer of DNA from donor to recipient cell, the genes are designated with the three-letter names of *tra* or *trb*, followed by a capital letter. The *tra* genes are shown in red, and the *trb* genes are shown in blue. The functions of a few examples are indicated. The origin of transfer is designated *oriT*.

Concept check: *Would this circular DNA molecule be found in an F^+ or F^- cell?*

The successful contact between a donor and recipient cell stimulates the donor cell to begin the transfer process. Genes within the F factor encode a protein complex called the **relaxosome.** This complex first recognizes a DNA sequence in the F factor known as the **origin of transfer** (see Figure 9.4a). Upon recognition, one DNA strand in the site is cut. The relaxosome also catalyzes the separation of the DNA strands, and only the cut DNA strand is transferred to the recipient cell. As the DNA strands separate, most of the proteins within the relaxosome are released, but a protein called relaxase remains bound to one end of the cut DNA strand.

The next phase of conjugation involves the export of the nucleoprotein complex from the donor cell to the recipient cell. To begin this process, the DNA/relaxase complex is recognized by a coupling factor that promotes the entry of the nucleoprotein into the exporter, a complex of proteins that spans both inner and outer membranes of the donor cell. In bacterial species, this complex is formed from 10 to 15 different proteins that are encoded by genes within the F factor.

Once the DNA/relaxase complex is pumped out of the donor cell, it travels through the conjugation bridge and then into the recipient cell. As shown in Figure 9.4a, the other strand of the F factor DNA remains in the donor cell, where DNA replication restores the F factor DNA to its original double-stranded condition. After the recipient cell receives a single strand of the F factor DNA, relaxase catalyzes the joining of the ends of the linear DNA molecule to form a circular molecule. This single-stranded DNA is replicated in the recipient cell to become double-stranded. The result of conjugation is that the recipient cell has acquired an F factor, converting it from an F^- to an F^+ cell. The genetic composition of the donor strain has not changed.

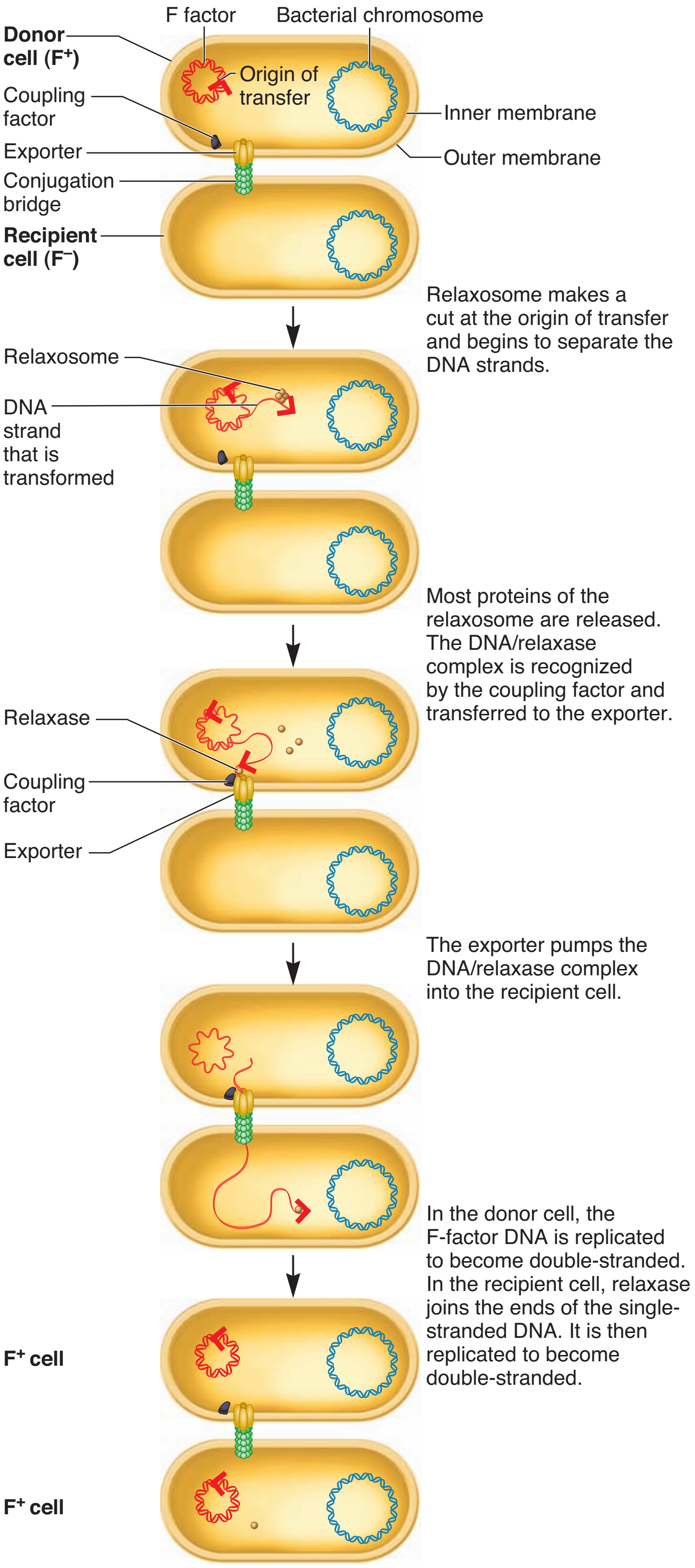
Bacteria May Contain Different Types of Plasmids

An F factor is one type of DNA that exists independently of the chromosomal DNA. The more general term for this type of structure is a **plasmid.** Most known plasmids are circular, although some are linear. Plasmids occur naturally in many strains of bacteria and in a few types of eukaryotic cells such as yeast. The smallest plasmids consist of just a few thousand base pairs (bp) and carry only a gene or two; the largest are in the range of 100,000 to 500,000 bp and carry several dozen or even hundreds of genes. Some plasmids, such as F factors, can integrate into the chromosome. These are also called **episomes.**

A plasmid has its own origin of replication that allows it to be replicated independently of the bacterial chromosome. The DNA sequence of the origin of replication influences how many copies of the plasmid are found within a cell. Some origins are said to be very strong because they result in many copies of the plasmid, perhaps as many as 100 per cell. Other origins of replication have sequences that are described as much weaker, in that the number of copies created is relatively low, such as one or two per cell.

Why do bacteria have plasmids? Plasmids are not usually necessary for bacterial survival. However, in many cases, certain genes within a plasmid provide some type of growth advantage to the cell. By studying plasmids in many different species, researchers have discovered that most plasmids fall into five different categories:

1. Fertility plasmids, also known as F factors, allow bacteria to mate with each other.
2. Resistance plasmids, also known as R factors, contain genes that confer resistance against antibiotics and other types of toxins.


F factor
Bacterial chromosome
Donor cell (F⁺)
Origin of transfer
Coupling factor
Inner membrane
Exporter
Outer membrane
Conjugation bridge
Recipient cell (F⁻)
Relaxosome makes a cut at the origin of transfer and begins to separate the DNA strands.
Relaxosome
DNA strand that is transformed
Most proteins of the relaxosome are released. The DNA/relaxase complex is recognized by the coupling factor and transferred to the exporter.
Relaxase
Coupling factor
Exporter
The exporter pumps the DNA/relaxase complex into the recipient cell.
In the donor cell, the F-factor DNA is replicated to become double-stranded. In the recipient cell, relaxase joins the ends of the single-stranded DNA. It is then replicated to become double-stranded.
F⁺ cell
F⁺ cell

(a) Transfer of an F factor via conjugation

(b) Conjugating *E. coli*

ONLINE ANIMATION **FIGURE 9.4** **The transfer of an F factor during bacterial conjugation.** **(a)** The mechanism of transfer. The end result is that both cells have an F factor. **(b)** Two *E. coli* cells in the act of conjugation. The cell on the left is F⁺, and the one on the right is F⁻. The two cells make contact with each other via sex pili that are made by the F⁺ cell.

Concept check: What are the functions of relaxase, coupling factor, and the exporter?

3. Degradative plasmids carry genes that enable the bacterium to digest and utilize an unusual substance. For example, a degradative plasmid may carry genes that allow a bacterium to digest an organic solvent such as toluene.
4. Col-plasmids contain genes that encode colicins, which are proteins that kill other bacteria.
5. Virulence plasmids carry genes that turn a bacterium into a pathogenic strain.

9.2 REVIEWING THE KEY CONCEPTS

- Joshua Lederberg and Edward Tatum discovered conjugation in *E. coli* by analyzing auxotrophic strains (see Figure 9.1).
- Using a U-tube apparatus, Davis showed that conjugation requires cell-to-cell contact (see Figure 9.2).
- Certain strains of bacteria have F factors, which they can transfer via conjugation in a series of steps (see Figures 9.3, 9.4).
- Bacteria may contain different types of plasmids.

9.2 COMPREHENSION QUESTIONS

1. A bacterial cell with an F factor is mated to an F⁻ cell. Following mating, the two cells would be
   a. F⁺.
   b. F⁻.
   c. one F⁺ and one F⁻.
   d. none of the above.
2. Which of the following is a type of plasmid?
   a. F factor (fertility factor)
   b. R factor (resistance plasmid)
   c. Virulence plasmids
   d. All of the above

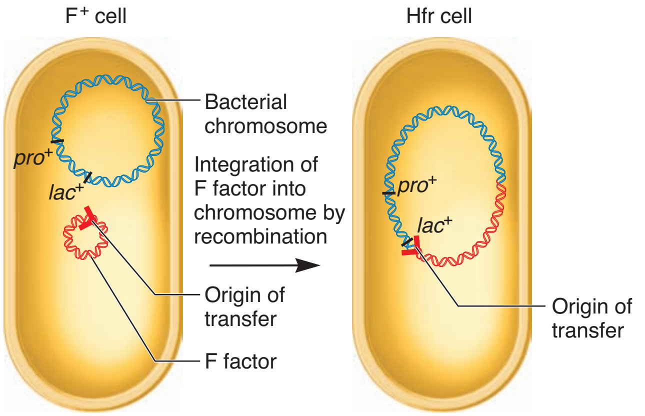
9.3 CONJUGATION AND MAPPING VIA *Hfr* STRAINS

Thus far, we have considered how conjugation may involve the transfer of an F factor from a donor to a recipient cell. In addition to this form of conjugation, other donor strains of *E. coli*, called ***Hfr* strains** (for high frequency of recombination), are capable of conjugation. In this section, we will examine how *Hfr* strains are formed and how they transfer genes to recipient cells. We will also explore the use of *Hfr* strains to map genes along the *E. coli* chromosome.

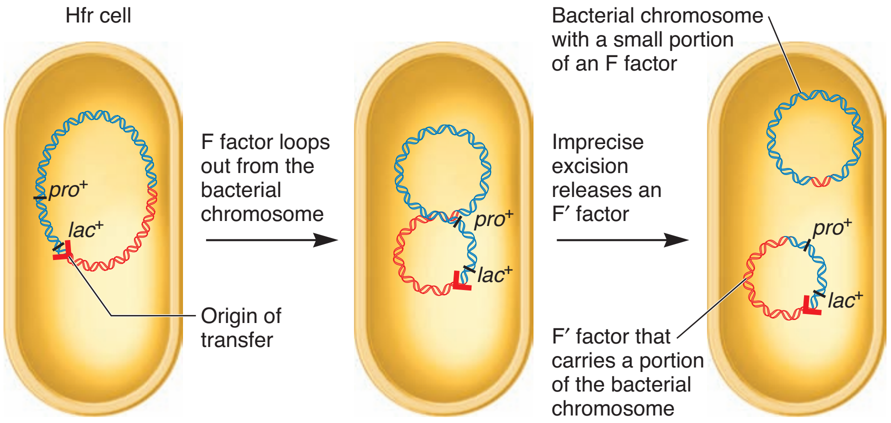
Hfr Strains Contain an F Factor Integrated into the Bacterial Chromosome

Luca Cavalli-Sforza discovered a strain of *E. coli* that was very efficient at transferring many chromosomal genes to recipient F^- strains. Cavalli-Sforza designated this bacterial strain an *Hfr* strain. How is an *Hfr* strain formed? As shown in **Figure 9.5a**, an F factor may align with a similar region found in the bacterial chromosome. Due to recombination, which is described in Chapter 13, the F factor may integrate into the bacterial chromosome. In this example, the F factor has integrated next to a *lac*$^+$ gene. F factors can integrate into several different sites that are scattered around the *E. coli* chromosome.

Occasionally, the integrated F factor in an *Hfr* strain is excised from the bacterial chromosome. This process involves the looping out of the F factor DNA from the chromosome, which is followed by recombination that releases the F factor from the chromosome (**Figure 9.5b**). In the example shown here, the excision is imprecise. This produces an F factor that carries a portion of the bacterial chromosome and leaves behind some of the F factor DNA in the bacterial chromosome. F factors that carry portions of the bacterial chromosome are called **F′ factors** (F prime factors). For example, in the experiment of Lederberg and Tatum described earlier in Figure 9.1, an F′ factor carrying the *bio*$^+$ and *met*$^+$ genes or an F′ factor carrying the *thr*$^+$, *leu*$^+$,


F$^+$ cell
Hfr cell
Bacterial chromosome
pro$^+$
lac$^+$
Integration of F factor into chromosome by recombination
Origin of transfer
F factor
pro$^+$
lac$^+$
Origin of transfer

(a) When an F factor integrates into the chromosome, it creates an Hfr cell.


Hfr cell
Bacterial chromosome with a small portion of an F factor
pro$^+$
lac$^+$
F factor loops out from the bacterial chromosome
Origin of transfer
pro$^+$
lac$^+$
Imprecise excision releases an F′ factor
pro$^+$
lac$^+$
F′ factor that carries a portion of the bacterial chromosome

(b) When an F factor excises imprecisely, an F′ factor is created.

FIGURE 9.5 **Integration of an F factor to form an Hfr cell and its subsequent excision to form an F′ factor. (a)** An Hfr cell is created when an F factor integrates into the bacterial chromosome. **(b)** When an F factor is imprecisely excised, an F′ factor is created that carries a portion of the bacterial chromosome.

Concept check: *How is an F′ factor different from an F factor?*

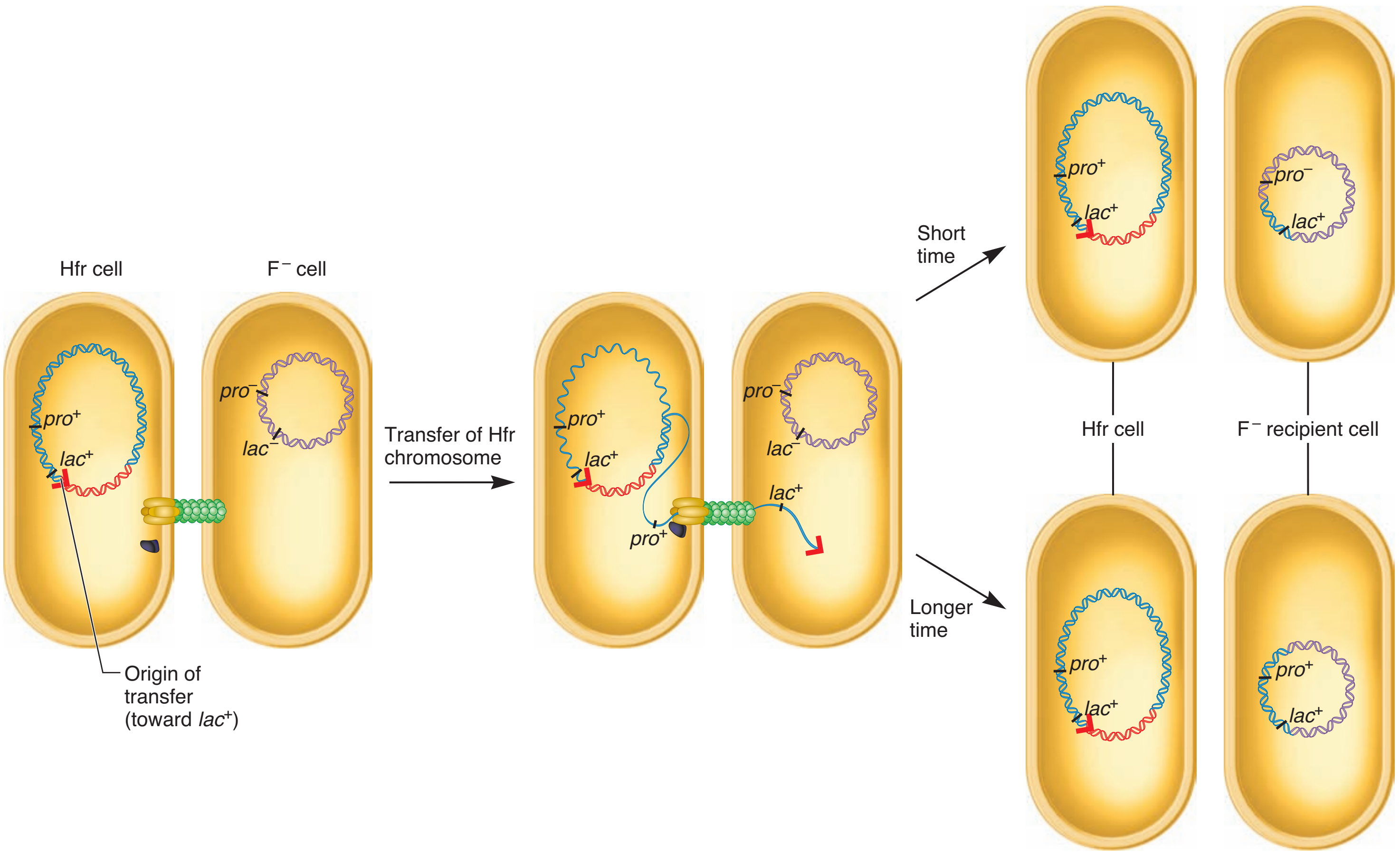


FIGURE 9.6 **Transfer of bacterial genes from an Hfr cell to an F^- cell.** The transfer of the bacterial chromosome begins at the origin of transfer and then proceeds around the circular chromosome. After a segment of chromosome has been transferred to the F^- recipient cell, it recombines with the recipient cell's chromosome. If mating occurs for a brief period, only a short segment of the chromosome is transferred. If mating is prolonged, a longer segment of the bacterial chromosome is transferred.

Genes → Traits The F^- recipient cell was originally *lac*$^-$ (unable to metabolize lactose) and *pro*$^-$ (unable to synthesize proline). If mating occurred for a short period of time, the recipient cell acquired *lac*$^+$, allowing it to metabolize lactose. If mating occurred for a longer period of time, the recipient cell also acquired *pro*$^+$, enabling it to synthesize proline.

Concept check: *With regard to the timing of conjugation, explain why the recipient cell in the top right is pro*$^-$ *whereas the recipient cell in the bottom right is pro*$^+$.

and *thi*$^+$ genes may have been transferred to the recipient strain. Therefore, conjugation may introduce new genes into the recipient strain, thereby altering its genotype.

Hfr Strains Can Transfer a Portion of the Bacterial Chromosome to Recipient Cells

William Hayes, who independently isolated another *Hfr* strain, demonstrated that conjugation between an Hfr cell and an F^- cell involves the transfer of a portion of the bacterial chromosome from the Hfr cell to the F^- cell (**Figure 9.6**). The origin of transfer within the integrated F factor determines the starting point and direction of this transfer process. One of the DNA strands is cut, or nicked, at the origin of transfer. This cut site is the starting point at which the Hfr chromosome enters the F^- recipient cell. From this starting point, a strand of the DNA of the Hfr chromosome begins to enter the F^- cell in a linear manner. The transfer process occurs in conjunction with chromosomal replication, so the Hfr cell retains its original chromosomal composition. About 1.5 to 2 hours is required for the entire Hfr chromosome to pass into the F^- cell. Because most matings do not last that long, usually only a portion of the Hfr chromosome is transmitted to the F^- cell.

Once inside the F^- cell, the chromosomal material from the Hfr cell can swap, or recombine, with the homologous region of the recipient cell's chromosome. (Chapter 13 describes the process of homologous recombination.) How does this process affect the recipient cell? As illustrated in Figure 9.6, this recombination may provide the recipient cell with a new combination of alleles. In this example, the recipient strain was originally *lac*$^-$ (unable to metabolize lactose) and *pro*$^-$ (unable to synthesize proline). If mating occurred for a short time, the recipient cell received a short segment of chromosomal DNA from the donor. In this case, the recipient cell has become *lac*$^+$ but remains *pro*$^-$. If the mating is prolonged, the recipient cell will receive a longer segment of chromosomal DNA from the donor. After a longer mating, the recipient becomes lac^{+} and pro^{+}. As shown in Figure 9.6, an important feature of Hfr mating is that the bacterial chromosome is transferred linearly to the recipient strain. In this example, lac^{+} is always transferred first, and pro^{+} is transferred later.

In any particular *Hfr* strain, the origin of transfer has a specific orientation that promotes either a counterclockwise or clockwise transfer of genes. Among different *Hfr* strains, the origin of transfer may be located in different regions of the chromosome. Therefore, the order of gene transfer depends on the location and orientation of the origin of transfer. For example, another *Hfr* strain could have its origin of transfer next to pro^{+} and transfer pro^{+} first and then lac^{+}.

Conjugation Experiments Can Map Genes Along the *E. coli* Chromosome

The first genetic mapping experiments in bacteria were carried out by Elie Wollman and François Jacob in the 1950s. At the time of their studies, not much was known about the organization of bacterial genes along the chromosome. A few key advances made Wollman and Jacob realize that the process of genetic transfer could be used to map the order of genes in *E. coli*. First, the discovery of conjugation by Joshua Lederberg and Edward Tatum and the identification of *Hfr* strains by Cavalli-Sforza and Hayes made it clear that bacteria can transfer genes from donor to recipient cells in a linear fashion. In addition, Wollman and Jacob were aware of previous microbiological studies concerning bacteriophages—viruses that bind to *E. coli* cells and subsequently infect them. These studies showed that bacteriophages can be sheared from the surface of *E. coli* cells if they are spun in a blender. In this treatment, the bacteriophages are detached from the surface of the bacterial cells, but the bacteria themselves remain healthy and viable. Wollman and Jacob reasoned that a blender treatment could also be used to separate bacterial cells that were in the act of conjugation without killing them. This technique is known as an **interrupted mating.**

The rationale behind Wollman and Jacob's mapping strategy is that the time it takes for genes to enter a donor cell is directly related to their order along the bacterial chromosome. They hypothesized that the chromosome of the donor strain in an *Hfr* mating is transferred in a linear manner to the recipient strain. If so, the order of genes along the chromosome can be deduced by determining the time it takes various genes to enter the recipient strain. Assuming the Hfr chromosome is transferred linearly, they realized that interruptions of mating at different times would lead to various lengths of the Hfr chromosome being transferred to the F^{-} recipient cell. If two bacterial cells had mated for a short period of time, only a small segment of the Hfr chromosome would be transferred to the recipient bacterium. However, if the bacterial cells were allowed to mate for a longer period before being interrupted, a longer segment of the Hfr chromosome could be transferred (see Figure 9.6). By determining which genes were transferred during short matings and which required longer times, Wollman and Jacob were able to deduce the order of particular genes along the *E. coli* chromosome.

As shown in the experiment of **Figure 9.7**, Wollman and Jacob began with two *E. coli* strains. The donor (*Hfr*) strain had the following genetic composition:

thr^{+}: able to synthesize threonine, an essential amino acid for growth
leu^{+}: able to synthesize leucine, an essential amino acid for growth
azi^{s}: sensitive to killing by azide (a toxic chemical)
ton^{s}: sensitive to infection by bacteriophage T1. (As discussed in Chapter 10, when bacteriophages infect bacteria, they may cause lysis, which results in plaque formation.)
lac^{+}: able to metabolize lactose and use it for growth
gal^{+}: able to metabolize galactose and use it for growth
str^{s}: sensitive to killing by streptomycin (an antibiotic)

The recipient (F^{-}) strain had the opposite genotype: thr^{-} leu^{-} azi^{r} ton^{r} lac^{-} gal^{-} str^{r} (r = resistant). Before the experiment, Wollman and Jacob already knew the thr^{+} gene was transferred first, followed by the leu^{+} gene, and both were transferred relatively soon (5–10 minutes) after mating. Their main goal in this experiment was to determine the times at which the other genes (azi^{s} ton^{s} lac^{+} gal^{+}) were transferred to the recipient strain. The transfer of the str^{s} gene was not examined because streptomycin was used to kill the donor strain following conjugation.

Before discussing the interpretation of this experiment, let's consider how Wollman and Jacob monitored gene transfer. To determine if particular genes had been transferred after mating, they took the mated cells and first plated them on a growth medium that lacked threonine (thr) and leucine (leu) but contained streptomycin (str). On these plates, the original donor and recipient strains could not grow because the donor strain was streptomycin-sensitive and the recipient strain required threonine and leucine. However, mated cells in which the donor had transferred chromosomal DNA carrying the thr^{+} and leu^{+} genes to the recipient cell would be able to grow.

To determine the order of gene transfer of the azi^{s}, ton^{s}, lac^{+}, and gal^{+} genes, Wollman and Jacob picked colonies from the first plates and restreaked them on media that contained azide or bacteriophage T1 or on media that contained lactose or galactose as the sole source of energy for growth. The plates were incubated overnight to observe the formation of visible bacterial growth. Whether or not the bacteria could grow depended on their genotypes. For example, a cell that is azi^{s} cannot grow on a medium containing azide, and a cell that is lac^{-} cannot grow on a medium containing lactose as the carbon source for growth. By comparison, a cell that is azi^{r} and lac^{+} can grow on both types of media.

▶ THE GOAL (DISCOVERY-BASED SCIENCE)

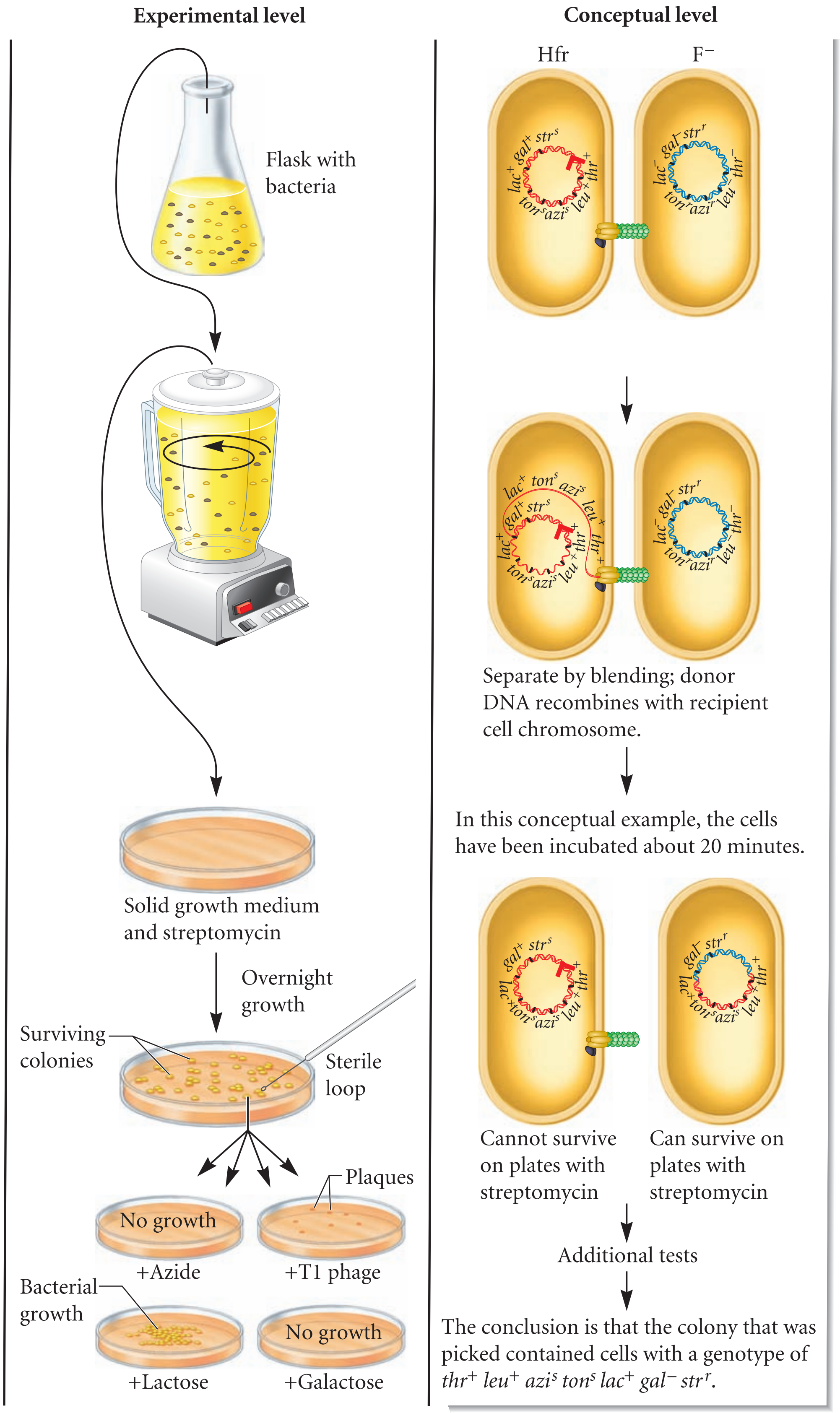
The chromosome of the donor strain in an *Hfr* mating is transferred in a linear manner to the recipient strain. The order of genes along the chromosome can be deduced by determining the time various genes take to enter the recipient strain.

ACHIEVING THE GOAL — FIGURE 9.7 The use of conjugation to map the order of genes along the *E. coli* chromosome.

Concept check: *Explain how the researchers achieved interrupted mating.*

Starting materials: The two *E. coli* strains already described, one *Hfr* strain (thr^{+} leu^{+} azi^{s} ton^{s} lac^{+} gal^{+} str^{s}) and one F^{-} (thr^{-} leu^{-} azi^{r} ton^{r} lac^{-} gal^{-} str^{r}).

1. Mix together a large number of Hfr donor and F⁻ recipient cells.
2. After different periods of time, take a sample of cells and interrupt conjugation in a blender.
3. Plate the cells on growth media lacking threonine and leucine but containing streptomycin. Note: The general methods for growing bacteria in a laboratory are described in the Appendix.
4. Pick each surviving colony, which would have to be thr^{+} leu^{+} str^{r}, and test to see if it is sensitive to killing by azide, sensitive to infection by T1 bacteriophage, and able to metabolize lactose or galactose.


Experimental level
Conceptual level
Flask with bacteria
Hfr
F⁻
lac⁺ gal⁺ str^s
ton^s azi^s leu⁺ thr⁺
lac⁻ gal⁻ str^r
ton^r azi^r leu⁻ thr⁻
Solid growth medium and streptomycin
Overnight growth
Surviving colonies
Sterile loop
Plaques
No growth
+Azide
+T1 phage
Bacterial growth
No growth
+Lactose
+Galactose
Additional tests

Separate by blending; donor DNA recombines with recipient cell chromosome.

In this conceptual example, the cells have been incubated about 20 minutes.

Cannot survive on plates with streptomycin

Can survive on plates with streptomycin

The conclusion is that the colony that was picked contained cells with a genotype of thr^{+} leu^{+} azi^{s} ton^{s} lac^{+} gal^{-} str^{r}.

▶ THE DATA

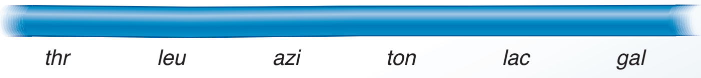
Minutes That Bacterial Cells Were Allowed to Mate Before Blender Treatment	*Percent of Surviving Bacterial Colonies with the Following Genotypes:*				
	thr$^+$ *leu*$^+$	*azi*^{s}	*ton*^{s}	*lac*$^+$	*gal*$^+$
5	—*	—	—	—	—
10	100	12	3	0	0
15	100	70	31	0	0
20	100	88	71	12	0
25	100	92	80	28	0.6
30	100	90	75	36	5
40	100	90	75	38	20
50	100	91	78	42	27
60	100	91	78	42	27

*There were no surviving colonies within the first 5 minutes of mating.

Data from François Jacob and Elie Wollman (1961) *Sexuality and the Genetics of Bacteria.* Academic Press, New York.

▶ INTERPRETING THE DATA

Now let's discuss the data shown in Figure 9.7. After the first plating, all survivors would be cells in which the *thr*$^+$ and *leu*$^+$ alleles had been transferred to the *F*$^-$ recipient strain, which was already streptomycin-resistant. As seen in the data, 5 minutes was not sufficient time to transfer the *thr*$^+$ and *leu*$^+$ alleles because no surviving colonies were observed. After 10 minutes or longer, however, surviving bacterial colonies with the *thr*$^+$ *leu*$^+$ genotype were obtained. To determine the order of the remaining genes (*azi*^{s}, *ton*^{s}, *lac*$^+$, and *gal*$^+$), each surviving colony was tested to see if it was sensitive to killing by azide, sensitive to infection by T1 bacteriophage, able to use lactose for growth, or able to use galactose for growth. The likelihood of surviving colonies depended on whether the *azi*^{s}, *ton*^{s}, *lac*$^+$, and *gal*$^+$ genes were close to the origin of transfer or farther away. For example, when cells were allowed to mate for 25 minutes, 80% carried the *ton*^{s} gene, whereas only 0.6% carried the *gal*$^+$ gene. These results indicate that the *ton*^{s} gene is closer to the origin of transfer compared with the *gal*$^+$ gene. When comparing the data in Figure 9.7, a consistent pattern emerged. The gene that conferred sensitivity to azide (*azi*^{s}) was transferred first, followed by *ton*^{s}, *lac*$^+$, and finally, *gal*$^+$. From these data, as well as those from other experiments, Wollman and Jacob constructed a genetic map that described the order of these genes along the *E. coli* chromosome.



This work provided the first method for bacterial geneticists to map the order of genes along the bacterial chromosome.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

A Genetic Map of the *E. coli* Chromosome Has Been Obtained from Many Conjugation Studies

Conjugation experiments have been used to map more than 1000 genes along the circular *E. coli* chromosome. A map of the *E. coli* chromosome is shown in **Figure 9.8**. This simplified map shows the locations of only a few dozen genes. Because the chromosome is circular, we must arbitrarily assign a starting point on the map, in this case the gene *thrA*. Researchers scale genetic maps from bacterial conjugation studies in units of **minutes.** This unit refers to the relative time it takes for genes to first enter an *F*$^-$ recipient strain during a conjugation experiment. The *E. coli* genetic map shown in Figure 9.8 is 100 minutes long, which is approximately the time that it takes to transfer the complete chromosome during an *Hfr* mating.

The distance between two genes is determined by comparing their times of entry during a conjugation experiment. As shown in **Figure 9.9**, the time of entry is found by conducting mating experiments at different time intervals before interruption. We compute the time of entry by extrapolating the data back to the *x*-axis. In this experiment, the time of entry of the *lacZ* gene was approximately 16 minutes, and that of the *galE* gene was 25 minutes. Therefore, these two genes are approximately 9 minutes apart from each other along the *E. coli* chromosome.

Let's look back at Figure 9.8 and consider where the origin of transfer must have been located in the donor strain that was used in

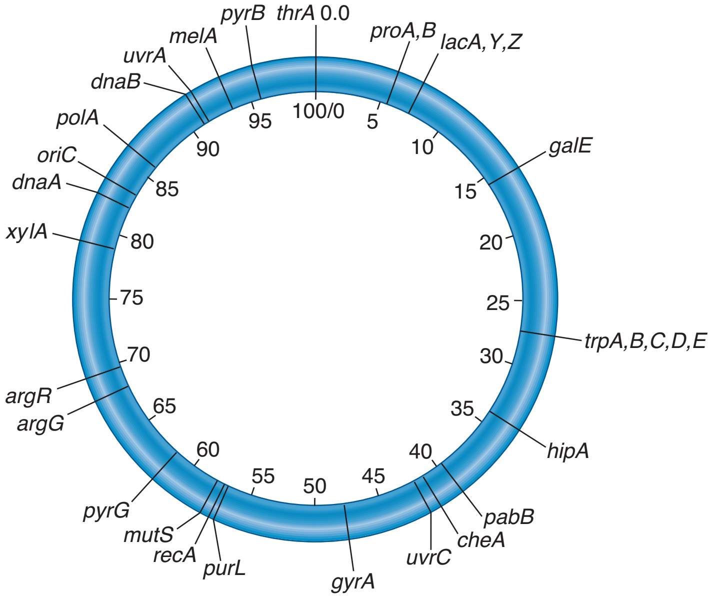


FIGURE 9.8 **A simplified genetic map of the *E. coli* chromosome indicating the positions of several genes.** *E. coli* has a circular chromosome with about 4377 different genes. This map shows the locations of several of them. The map is scaled in units of minutes and proceeds in a clockwise direction. The starting point on the map is the gene *thrA*.

Concept check: *Why is the scale of this map in minutes?*

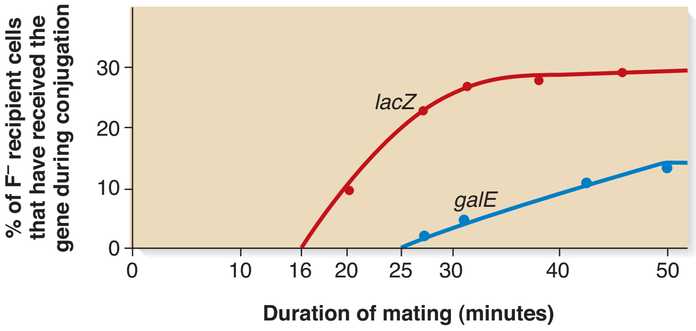


FIGURE 9.9 **Time course of an interrupted *E. coli* conjugation experiment.** By extrapolating the data back to the origin, the approximate time of entry of the *lacZ* gene is found to be 16 minutes; that of the *galE* gene is 25 minutes. Therefore, the distance between these two genes is 9 minutes.

Concept check: Which of these two genes is closer to the origin of transfer?

the experiment of Figure 9.9. The *lacZ* gene is located at 7 minutes on the chromosome map, and the *galE* is found at 16 minutes. For the donor strain used in Figure 9.9, we can deduce that the origin of transfer was located at approximately 91 minutes on the chromosome map and transferred DNA in the clockwise direction. When the origin is located at 91 minutes, it would take about 16 minutes to transfer *lacZ* and about 25 minutes to transfer *galE*.

9.3 REVIEWING THE KEY CONCEPTS

- *Hfr* strains are formed when an F factor integrates into the bacterial chromosome. An imprecise excision produces an F′ factor that carries a portion of the bacterial chromosome (see Figure 9.5).
- *Hfr* strains can transfer a portion of the bacterial chromosome to a recipient cell during conjugation (see Figure 9.6).
- Wollman and Jacob showed that conjugation can be used to map the locations of genes along the bacterial chromosome, thereby creating a genetic map (see Figures 9.7–9.9).

9.3 COMPREHENSION QUESTIONS

1. With regard to conjugation, a key difference between F $^+$ and Hfr cells is that an Hfr cell
   a. is unable to conjugate.
   b. transfers a plasmid to the recipient cell.
   c. transfers a portion of the bacterial chromosome to the recipient cell.
   d. becomes an F$^-$ cell after conjugation.
2. In mapping experiments, ______ strains are mated to *F*$^-$ strains. The distance between two genes is determined by comparing their __________________ during a conjugation experiment.
   a. *F*$^+$, times of entry
   b. *Hfr*, times of entry
   c. *F*$^+$, expression levels
   d. *Hfr*, expression levels

9.4 BACTERIAL TRANSDUCTION

We now turn to a second method of genetic transfer, one that involves bacteriophages, or phages, which are viruses that infect bacterial cells. Following infection of the cell, new viral particles are made, which are then released from the cell in an event called lysis. In this section, we will examine how mistakes can happen in the bacteriophage reproductive cycle that lead to the transfer of genetic material from one bacterial cell to another.

Bacteriophages Transfer Genetic Material from One Bacterial Cell to Another Via Transduction

The ability of phages to transfer genetic material between bacteria is called transduction. Examples of phages that can transfer bacterial chromosomal DNA from one bacterium to another are the P22 and P1 phages, which infect the bacterial species *Salmonella typhimurium* and *E. coli*, respectively. The P22 and P1 phages can follow either the lytic or lysogenic cycle. (Look ahead to Chapter 10, Figure 10.3 for a description of these two cycles.) In the lytic cycle, when the phage infects the bacterial cell, the bacterial chromosome becomes fragmented into small pieces of DNA. The phage DNA directs the synthesis of more phage DNA and proteins, which then assemble to make new phages and the bacterial DNA is degraded.

How does a bacteriophage transfer bacterial chromosomal genes from one cell to another? Occasionally, a mistake can happen in which a piece of bacterial DNA assembles with phage proteins (**Figure 9.10**). This creates a transducing phage that contains bacterial chromosomal DNA. When phage synthesis is completed, the bacterial cell is lysed and releases the newly made phages into the environment. Following release, the transducing phage can bind to a living bacterial cell and inject its genetic material into the bacterium. The DNA fragment, which was derived from the chromosomal DNA of the first bacterium, can then recombine with the recipient cell's chromosome. In this case, the recipient bacterium has been changed from a cell that was *his*$^-$ *lys*$^-$ (unable to synthesize histidine and lysine) to a cell that is *his*$^+$ *lys*$^-$ (able to synthesize histidine but unable to synthesize lysine). In the example shown in Figure 9.10, any piece of the bacterial chromosomal DNA can be incorporated into the phage. This type of transduction is called **generalized transduction.** By comparison, some phages carry out specialized transduction in which only particular bacterial genes are transferred to recipient cells (see Solved Problem S5 at the end of the chapter).

Transduction was first discovered in 1952 by Joshua Lederberg and Norton Zinder, using an experimental strategy similar to that depicted in Figure 9.1. They mixed together two strains of the bacterium *Salmonella typhimurium*. One strain, designated LA-22, was *phe*$^-$ *trp*$^-$ *met*$^+$ *his*$^+$. This strain was unable to synthesize phenylalanine or tryptophan but was able to synthesize methionine and histidine. The other strain, LA-2, was *phe*$^+$ *trp*$^+$ *met*$^-$ *his*$^-$. It was able to synthesize phenylalanine and tryptophan but not methionine or histidine. When a mixture of these cells was placed on plates with growth media lacking these four amino acids, approximately 1 cell in 100,000 was observed to grow. The genotype of the surviving bacterial cells must have been

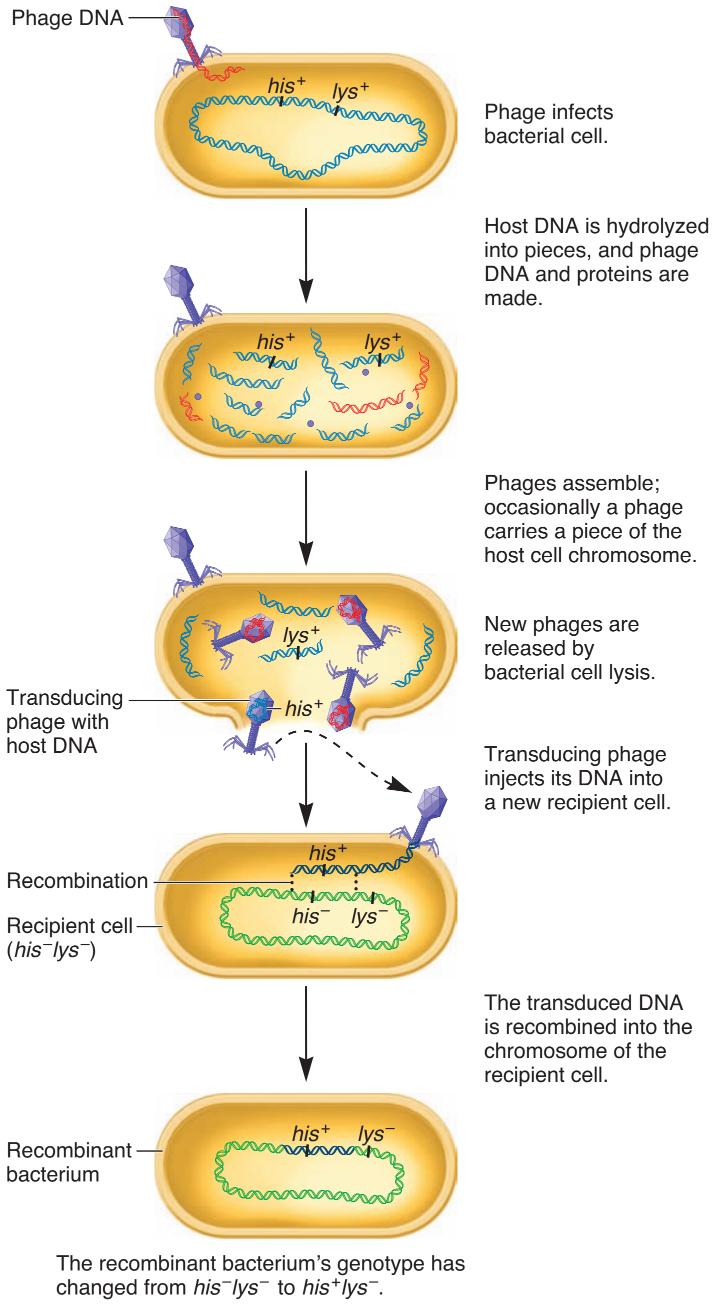

Phage DNA
his^+
lys^+
Phage infects bacterial cell.
Host DNA is hydrolyzed into pieces, and phage DNA and proteins are made.
his^+
lys^+
Phages assemble; occasionally a phage carries a piece of the host cell chromosome.
lys^+
New phages are released by bacterial cell lysis.
Transducing phage with host DNA
his^+
Transducing phage injects its DNA into a new recipient cell.
his^+
Recombination
Recipient cell (his^-lys^-)
his^-
lys^-
The transduced DNA is recombined into the chromosome of the recipient cell.
Recombinant bacterium
his^+
lys^-
The recombinant bacterium's genotype has changed from his^-lys^- to his^+lys^-.

FIGURE 9.10 Transduction in bacteria.

Genes → Traits The transducing phage introduced DNA into a new recipient cell that was originally his^- lys^- (unable to synthesize histidine and lysine). During transduction, it received a segment of bacterial chromosomal DNA that carried his^+. Following recombination, the recipient cell's genotype was changed to his^+, so it now could synthesize histidine.

Concept check: Transduction is sometimes described as a mistake in the bacteriophage reproductive cycle. Explain how it is a mistake.

phe^+ trp^+ met^+ his^+. Therefore, Lederberg and Zinder concluded that genetic material had been transferred between the two strains.

A novel result occurred when Lederberg and Zinder repeated this experiment using a U-tube apparatus, as previously shown in Figure 9.2. They placed the LA-22 strain (phe^- trp^- met^+ his^+) on one side of the filter and LA-2 (phe^+ trp^+ met^- his^-) on the other. After an incubation period of several minutes, they removed samples from either side of the tube and plated the cells on media lacking the four amino acids. Surprisingly, they obtained colonies from the side of the tube that contained LA-22 but not from the side that contained LA-2. From these results, they concluded that some agent, which was small enough to pass through the filter, was produced on the side of the tube containing LA-2, traversed the filter, and then was taken up by the LA-22 strain, converting it to a phe^+ trp^+ met^+ his^+ genotype. By conducting this type of experiment using filters with different pore sizes, Lederberg and Zinder found that the agent was slightly less than 0.1 μm in diameter, a size much smaller than a bacterium. They correctly concluded that the filterable agent in these experiments was a bacteriophage. In this case, the LA-2 strain contained a prophage, a phage such as P22 that is integrated into the chromosome during the lysogenic cycle. On occasion, the prophage switched to the lytic cycle and packaged a segment of bacterial DNA carrying the phe^+ and trp^+ genes. This phage then traversed the filter and injected the genes into LA-22.

Cotransduction Can Be Used to Map Genes That Are Within 2 Minutes of Each Other

Can transduction be used to map the distance between bacterial genes? The answer is yes, but only if the genes are relatively close together. During transduction, P1 phages cannot package pieces that are greater than 2% to 2.5% of the entire *E. coli* chromosome, and P22 phages cannot package pieces that are greater than 1% of the length of the *S. typhimurium* chromosome. If two genes are close together along the chromosome, a bacteriophage may package a single piece of the chromosome that carries both genes and transfer that piece to another bacterium. This phenomenon is called **cotransduction.** The likelihood that two genes will be cotransduced depends on how close together they lie. If two genes are far apart along a bacterial chromosome, they will never be cotransduced because the bacteriophage cannot physically package a DNA fragment that is larger than 1% to 2.5% of the bacterial chromosome. In genetic mapping studies, cotransduction is used to determine the order and distance between genes that lie fairly close to each other.

To map genes using cotransduction, a researcher selects for the transduction of one gene and then monitors whether or not a second gene is cotransduced along with it. As an example, let's consider a donor strain of *E. coli* that is arg^+ met^+ str^s (able to synthesize arginine and methionine but sensitive to killing by streptomycin) and a recipient strain that is arg^- met^- str^r (Figure 9.11). The donor strain is infected with phage P1. Some of the *E. coli* cells are lysed by P1, and this P1 lysate is mixed with the recipient cells. After allowing sufficient time for transduction, the recipient cells are plated on a growth medium that contains arginine and streptomycin but not methionine. Therefore, these plates select for the growth of cells in which the met^+ gene has been transferred to the recipient strain, but they do not select for the growth of cells in which the arg^+ gene has been transferred, because the growth media are supplied with arginine.

Nevertheless, the arg^+ gene may have been cotransduced with the met^+ gene if the two genes are close together. To determine this, a sample of cells from each bacterial colony can be picked up with a wire loop and restreaked on media that lack both amino acids. If the colony can grow, the cells must have also obtained the arg^+ gene during transduction. In other words, cotransduction of both the arg^+ and met^+ genes has occurred. Alternatively, if the restreaked colony does not grow, it must have received only the met^+ gene during transduction. Data from this type of experiment are shown at the bottom of Figure 9.11. These data indicate a cotransduction frequency of 21/50 = 0.42, or 42%.

In 1966, Tai Te Wu derived a mathematical expression that relates cotransduction frequency with map distance obtained from conjugation experiments. This equation is

$$\text{Cotransduction frequency} = (1 - d/L)^3$$

where

d = distance between two genes in minutes
L = the size of the chromosomal pieces (in minutes) that the phage carries during transduction (For P1 transduction, this size is approximately 2% of the bacterial chromosome, which equals about 2 minutes.)

This equation assumes the bacteriophage randomly packages pieces of the bacterial chromosome that are similar in size. Depending on the type of phage used in a transduction experiment, this assumption may not always be valid. Nevertheless, this equation has been fairly reliable in computing map distance for P1 transduction experiments in *E. coli*. We can use this equation to estimate the distance between the two genes described in Figure 9.11.

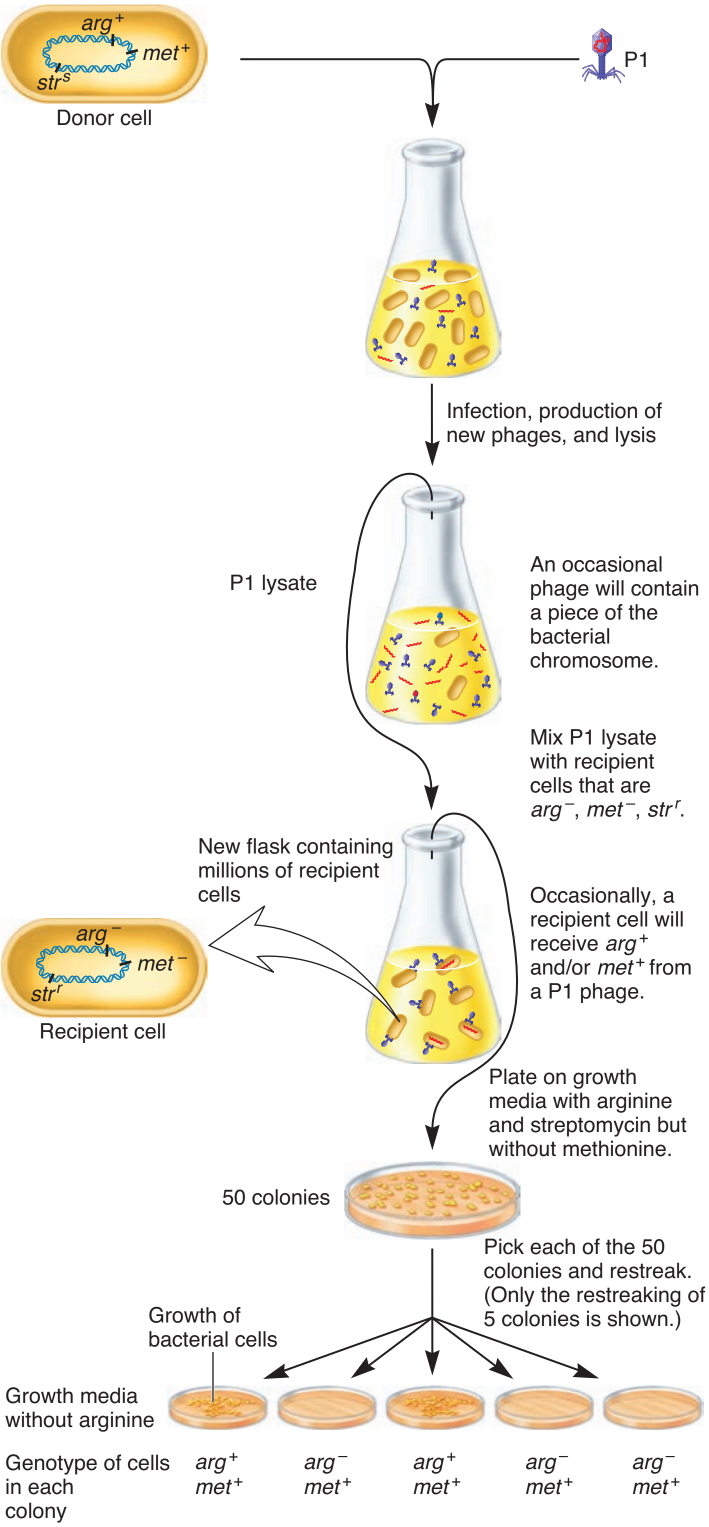
$$0.42 = (1 - d/2)^3$$

$$(1 - d/2) = \sqrt[3]{0.42}$$

$$1 - d/2 = 0.75$$

$$d/2 = 0.25$$

$$d = 0.5 \text{ minutes}$$


arg^+
met^+
str^s
Donor cell
P1
Infection, production of new phages, and lysis
P1 lysate
An occasional phage will contain a piece of the bacterial chromosome.
Mix P1 lysate with recipient cells that are arg^-, met^-, str^r.
New flask containing millions of recipient cells
arg^-
met^-
str^r
Recipient cell
Occasionally, a recipient cell will receive arg^+ and/or met^+ from a P1 phage.
Plate on growth media with arginine and streptomycin but without methionine.
50 colonies
Pick each of the 50 colonies and restreak. (Only the restreaking of 5 colonies is shown.)
Growth of bacterial cells
Growth media without arginine
Genotype of cells in each colony
arg^+ met^+
arg^- met^+
arg^+ met^+
arg^- met^+
arg^- met^+

Results

Selected gene	Nonselected gene	Number of colonies that grew on media + arginine	media – arginine	Cotransduction frequency
met^+	arg^+	50	21	0.42

FIGURE 9.11 **The steps in a cotransduction experiment.** The donor strain, which is arg^+ met^+ str^s (able to synthesize arginine and methionine but sensitive to streptomycin), is infected with phage P1. Some of the cells are lysed by P1, and this P1 lysate is mixed with cells of the recipient strain, which are arg^- met^- str^r. P1 phages in this lysate may carry fragments of the donor cell's chromosome, and the P1 phage may inject that DNA into the recipient cells. To identify recipient cells that have received the met^+ gene from the donor strain, the recipient cells are plated on a growth medium that contains arginine and streptomycin but not methionine. To determine if the arg^+ gene has also been cotransduced, cells from each bacterial colony (on the plates containing arginine and streptomycin) are lifted with a sterile wire loop and streaked on a medium that lacks both amino acids. If the cells can grow, cotransduction of the arg^+ and met^+ genes has occurred. Alternatively, if the restreaked colony does not grow, it must have received only the met^+ gene during transduction.

Concept check: *If these two genes were very far apart on the bacterial chromosome, how would the results have been different?*

This equation tells us that the distance between the *met*$^+$ and *arg*$^+$ genes is approximately 0.5 minutes.

Historically, genetic mapping strategies in bacteria often involved data from both conjugation and transduction experiments. Conjugation has been used to determine the relative order and distance of genes, particularly those that are far apart along the chromosome. In comparison, transduction experiments can provide fairly accurate mapping data for genes that are close together.

9.4 REVIEWING THE KEY CONCEPTS

- During transduction, a portion of a bacterial chromosome is transferred to a recipient cell via a bacteriophage (see Figure 9.10).
- A cotransduction experiment can be used to map genes that are close together along a bacterial chromosome (see Figure 9.11).

9.4 COMPREHENSION QUESTIONS

1. The term *generalized transduction* refers to the phenomenon that
    a. any small fragment of the bacterial chromosome may be transferred to another bacterium by a bacteriophage.
    b. only a specific fragment of DNA may be transferred to another bacterium by a bacteriophage.
    c. any small fragment of the bacterial chromosome may be transferred during conjugation.
    d. only a specific fragment of DNA may be transferred during conjugation.
2. Cotransduction may be used to map bacterial genes that are
    a. far apart on the bacterial chromosome.
    b. close together on the bacterial chromosome.
    c. both b and c.
    d. neither b or c.

9.5 BACTERIAL TRANSFORMATION

A third mechanism for the transfer of genetic material from one bacterium to another is known as transformation. This process was first discovered by Frederick Griffith in 1928 while working with strains of *Streptococcus pneumoniae* (formerly known as *Diplococcus pneumoniae,* or pneumococcus). During transformation, a living bacterial cell takes up DNA that is released from a dead bacterium. This DNA may then recombine into the living bacterium's chromosome, producing a bacterium with genetic material that it has received from the dead bacterium. (This experiment is discussed in detail in Chapter 11, Figure 11.1.) Transformation may be either a natural process that has evolved in certain bacteria, in which case it is called **natural transformation,** or an artificial process in which the bacterial cells are forced to take up DNA, an experimental approach termed **artificial transformation.** For example, a technique known as electroporation, in which an electric current causes the uptake of DNA, is used by researchers to promote the transport of DNA into a bacterial cell.

Since the initial studies of Griffith, we have learned a great deal about the events that occur in natural transformation. This form of genetic transfer has been reported in a wide variety of bacterial species. Bacterial cells that are able to take up DNA are known as **competent cells.** Those that can take up DNA naturally carry genes that encode proteins called **competence factors.** These proteins facilitate the binding of DNA fragments to the cell surface, the uptake of DNA into the cytoplasm, and its subsequent incorporation into the bacterial chromosome. Temperature, ionic conditions, and nutrient availability can affect whether or not a bacterium is competent to take up genetic material from its environment. These conditions influence the expression of competence genes.

In recent years, geneticists have unraveled some of the steps that occur when competent bacterial cells are transformed by genetic material in their environment. **Figure 9.12** describes the steps of transformation. First, a large fragment of DNA binds to the surface of the bacterial cell. Competent cells express DNA receptors that promote such binding. Before entering the cell, however, this large piece of chromosomal DNA must be cut into smaller fragments. This cutting is accomplished by an extracellular bacterial enzyme known as an endonuclease, which makes occasional random cuts in the long piece of chromosomal DNA. At this stage, the DNA fragments are composed of double-stranded DNA.

In the next step, the DNA fragment begins its entry into the bacterial cytoplasm. For this to occur, the double-stranded DNA interacts with proteins in the bacterial membrane. One of the DNA strands is degraded, and the other strand enters the bacterial cytoplasm via an uptake system, which is structurally similar to the one described for conjugation (as shown earlier in Figure 9.4a), but is involved with DNA uptake rather than export.

To be stably inherited, the DNA strand must be incorporated into the bacterial chromosome. If the DNA strand has a sequence that is similar to a region of DNA in the bacterial chromosome, the DNA may be incorporated into the chromosome by a process known as **homologous recombination,** discussed in detail in Chapter 13. For this to occur, the single-stranded DNA aligns itself with the homologous location on the bacterial chromosome. In the example shown in Figure 9.12, the foreign DNA carries a functional *lys*$^+$ gene that aligns itself with a nonfunctional (mutant) *lys*$^-$ gene already present within the bacterial chromosome. The foreign DNA then recombines with one of the strands in the bacterial chromosome of the competent cell. In other words, the foreign DNA replaces one of the chromosomal strands of DNA, which is subsequently degraded. During homologous recombination, alignment of the *lys*$^-$ and the *lys*$^+$ alleles results in a region of DNA called a **heteroduplex** that contains one or more base sequence mismatches. However, the heteroduplex exists only temporarily. DNA repair enzymes in the recipient cell recognize the heteroduplex and repair it. In this example, the heteroduplex has been repaired by eliminating the mutation that caused the *lys*$^-$ genotype, thereby creating a *lys*$^+$ gene. Therefore, the recipient cell has been transformed from a *lys*$^-$ strain to a *lys*$^+$ strain. Alternatively, a DNA fragment that has entered a cell may not be homologous to any genes that are already found in the bacterial chromosome. In this case, the DNA strand may be incorporated at a random site in the chromosome. This process is known as **nonhomologous recombination** or **illegitimate recombination.**

Some bacteria preferentially take up DNA fragments from other bacteria of the same species or closely related species. How

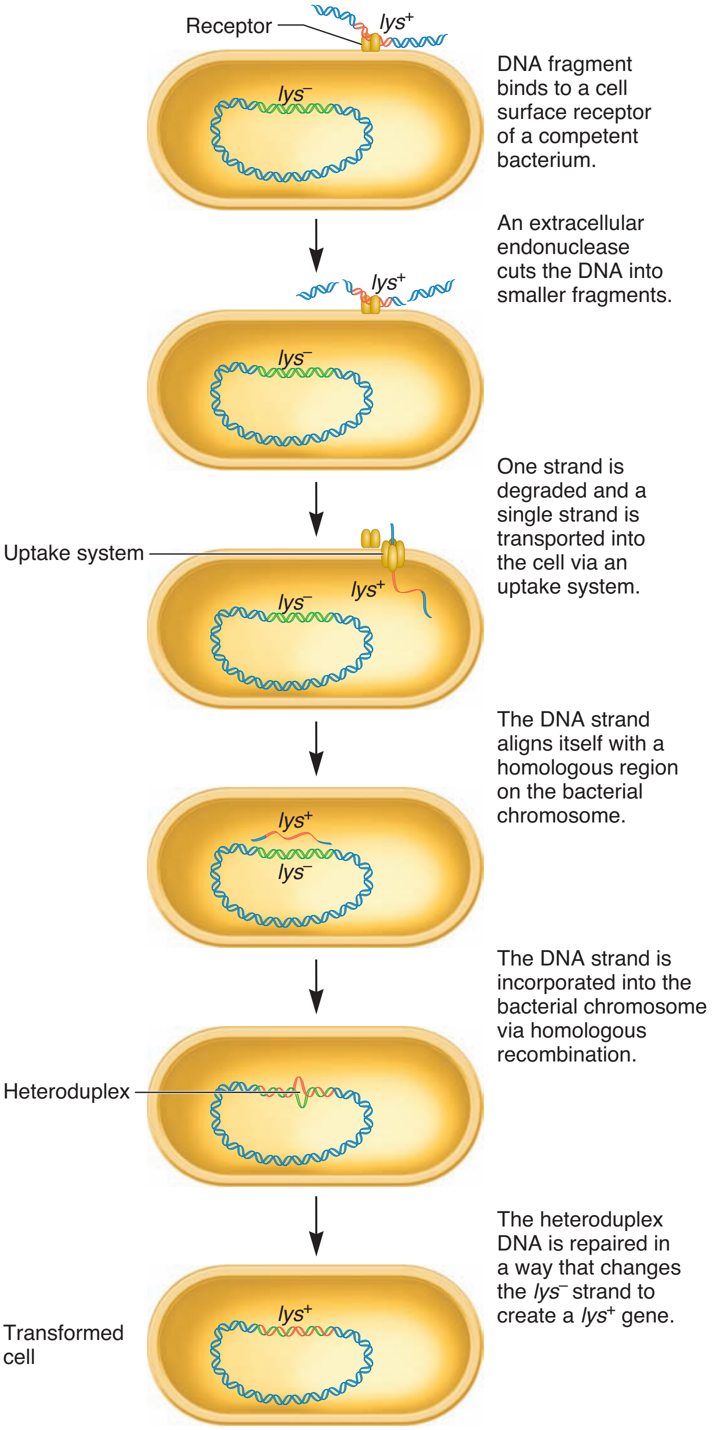

Receptor
lys+
lys−
DNA fragment binds to a cell surface receptor of a competent bacterium.
An extracellular endonuclease cuts the DNA into smaller fragments.
lys+
lys−
One strand is degraded and a single strand is transported into the cell via an uptake system.
Uptake system
lys−
lys+
The DNA strand aligns itself with a homologous region on the bacterial chromosome.
lys+
lys−
The DNA strand is incorporated into the bacterial chromosome via homologous recombination.
Heteroduplex
The heteroduplex DNA is repaired in a way that changes the lys− strand to create a lys+ gene.
Transformed cell
lys+

FIGURE 9.12 **The steps of bacterial transformation.** In this example, a fragment of DNA carrying a *lys*$^{+}$ gene enters the competent cell and recombines with the chromosome, transforming the bacterium from *lys*$^{-}$ to *lys*$^{+}$.

Genes → Traits Bacterial transformation can also lead to new traits for the recipient cell. The recipient cell was *lys*$^{-}$ (unable to synthesize the amino acid lysine). Following transformation, it became *lys*$^{+}$. The recipient bacterial cell was transformed into a cell that can synthesize lysine and grow on a medium that lacked this amino acid. Before transformation, the recipient *lys*$^{-}$ cell would not have been able to grow on a medium lacking lysine.

Concept check: *If the recipient cell did not already have a lys*$^{-}$ *gene, could the lys*$^{+}$ *DNA become incorporated into the bacterial chromosome? Explain.*

does this occur? Recent research has shown that the mechanism can vary among different species. In *Streptococcus pneumoniae*, the cells secrete a short peptide called the **competence-stimulating peptide (CSP).** When many *S. pneumoniae* cells are in the vicinity of one another, the concentration of CSP becomes high, which stimulates the cells, via a cell-signaling pathway, to express the competence proteins needed for the uptake of DNA and its incorporation in the chromosome. Because competence requires a high external concentration of CSP, *S. pneumoniae* cells are more likely to take up DNA from nearby *S. pneumoniae* cells that have died and released their DNA into the environment.

Other bacterial species promote the uptake of DNA among members of their own species via **DNA uptake signal sequences,** which are 9 or 10 bp long. In the human pathogens *Neisseria meningitidis* (a causative agent of meningitis), *N. gonorrhoeae* (a causative agent of gonorrhea), and *Haemophilus influenzae* (a causative agent of ear, sinus, and respiratory infections), these sequences are found at many locations within their respective genomes. For example, *H. influenzae* contains approximately 1500 copies of the sequence 5′-AAGTGCGGT-3′ in its genome, and *N. meningitidis* contains about 1900 copies of the sequence 5′-GCGTCTGAA-3′. DNA fragments that contain their own uptake signal sequence are preferentially taken up by these species instead of other DNA fragments. For example, *H. influenzae* is much more likely to take up a DNA fragment with the sequence 5′-AAGTGCGGT-3′. For this reason, transformation is more likely to involve DNA uptake between members of the same species.

Transformation has also been used to map many bacterial genes, using methods similar to the cotransduction experiments described earlier. If two genes are close together, the **cotransformation** frequency is expected to be high, whereas genes that are far apart have a cotransformation frequency that is very low or even zero. Like cotransduction, genetic mapping via cotransformation is used only to map genes that are relatively close together. Solved Problem S4 at the end of the chapter considers a cotransformation experiment.

9.5 REVIEWING THE KEY CONCEPTS

- During transformation, a segment of DNA is taken up by a bacterial cell and then is incorporated into the bacterial chromosome (Figure 9.12).

9.5 COMPREHENSION QUESTIONS

1. Some bacterial species preferentially take up DNA fragments from members of their own species. This can occur via
   a. competence-stimulating peptide (CSP).
   b. DNA uptake signal sequences.
   c. both a and b.
   d. none of the above.

KEY TERMS

Page 186. genetic transfer, conjugation, transduction, transformation
Page 187. minimal medium, auxotroph, prototroph
Page 188. F factor, sex pilus, conjugation bridge
Page 189. relaxosome, origin of transfer, plasmid, episomes
Page 191. *Hfr* strains, F′ factors
Page 193. interrupted mating
Page 195. minutes
Page 196. generalized transduction
Page 197. cotransduction
Page 199. natural transformation, artificial transformation, competent cells, competence factors, homologous recombination, heteroduplex, nonhomologous recombination, illegitimate recombination
Page 200. competence-stimulating peptide (CSP), DNA uptake signal sequences, cotransformation

CHAPTER SUMMARY

9.1 Overview of Genetic Transfer in Bacteria

- Three general mechanisms for genetic transfer in various species of bacteria are conjugation, transduction, and transformation (see Table 9.1).

9.2 Bacterial Conjugation

- Joshua Lederberg and Edward Tatum discovered conjugation in *E. coli* by analyzing auxotrophic strains (see Figure 9.1).
- Using a U-tube apparatus, Davis showed that conjugation requires cell-to-cell contact (see Figure 9.2).
- Certain strains of bacteria have F factors, which they can transfer via conjugation in a series of steps (see Figures 9.3, 9.4).
- Bacteria may contain different types of plasmids.

9.3 Conjugation and Mapping via *Hfr* Strains

- *Hfr* strains are formed when an F factor integrates into the bacterial chromosome. An imprecise excision produces an F′ factor that carries a portion of the bacterial chromosome (see Figure 9.5).
- *Hfr* strains can transfer a portion of the bacterial chromosome to a recipient cell during conjugation (see Figure 9.6).
- Wollman and Jacob showed that conjugation can be used to map the locations of genes along the bacterial chromosome, thereby creating a genetic map (see Figures 9.7–9.9).

9.4 Bacterial Transduction

- During transduction, a portion of a bacterial chromosome is transferred to a recipient cell via a bacteriophage (see Figure 9.10).
- A cotransduction experiment can be used to map genes that are close together along a bacterial chromosome (see Figure 9.11).

9.5 Bacterial Transformation

- During transformation, a segment of DNA is taken up by a bacterial cell and then is incorporated into the bacterial chromosome (Figure 9.12).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. In *E. coli*, the gene $bioD^+$ encodes an enzyme involved in biotin synthesis, and $galK^+$encodes an enzyme involved in galactose utilization. An *E. coli* strain that contained wild-type versions of both genes was infected with P1, and then a P1 lysate was obtained. This lysate was used to transduce (infect) a strain that was $bioD^-$ and $galK^-$. The cells were plated on media containing galactose as the sole carbon source for growth to select for transduction of the $galK^+$ gene. These media also were supplemented with biotin. The colonies were then restreaked on media that lacked biotin to see if the $bioD^+$ gene had been cotransduced. The following results were obtained:

		Number of Colonies That Grew on:		
Selected Gene	**Nonselected Gene**	**Galactose + Biotin**	**Galactose − Biotin**	**Cotransduction Frequency**
$galK^+$	$bioD^+$	80	10	0.125

How far apart are these two genes?

Answer: We can use the cotransduction frequency to calculate the distance between the two genes (in minutes) using the equation

$$\text{Cotransduction frequency} = (1 - d/2)^3$$
$$0.125 = (1 - d/2)^3$$
$$1 - d/2 = \sqrt[3]{0.125}$$
$$1 - d/2 = 0.5$$
$$d/2 = 1 - 0.5$$
$$d = 1.0 \text{ minute}$$

The two genes are approximately 1 minute apart on the *E. coli* chromosome.

S2. By conducting mating experiments between *Hfr* and recipient strains, Wollman and Jacob mapped the order of many bacterial genes. Throughout the course of their studies, they identified several different *Hfr* strains in which the F factor DNA had been integrated at different places along the bacterial chromosome. A sample of their experimental results is shown in the following table:

		Order of Transfer of Several Different Bacterial Genes								
***Hfr* strain**	**Origin**	**First**								**Last**
H	O	*thr*	*leu*	*azi*	*ton*	*pro*	*lac*	*gal*	*str*	*met*
1	O	*leu*	*thr*	*met*	*str*	*gal*	*lac*	*pro*	*ton*	*azi*
2	O	*pro*	*ton*	*azi*	*leu*	*thr*	*met*	*str*	*gal*	*lac*
3	O	*lac*	*pro*	*ton*	*azi*	*leu*	*thr*	*met*	*str*	*gal*
4	O	*met*	*str*	*gal*	*lac*	*pro*	*ton*	*azi*	*leu*	*thr*
5	O	*met*	*thr*	*leu*	*azi*	*ton*	*pro*	*lac*	*gal*	*str*
6	O	*met*	*thr*	*leu*	*azi*	*ton*	*pro*	*lac*	*gal*	*str*
7	O	*ton*	*azi*	*leu*	*thr*	*met*	*str*	*gal*	*lac*	*pro*

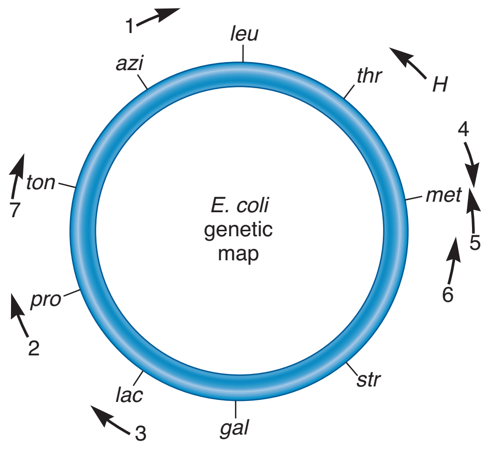
A. Explain how these results are consistent with the idea that the bacterial chromosome is circular.

B. Draw a map that shows the order of genes and the locations of the origins of transfer among these different *Hfr* strains.

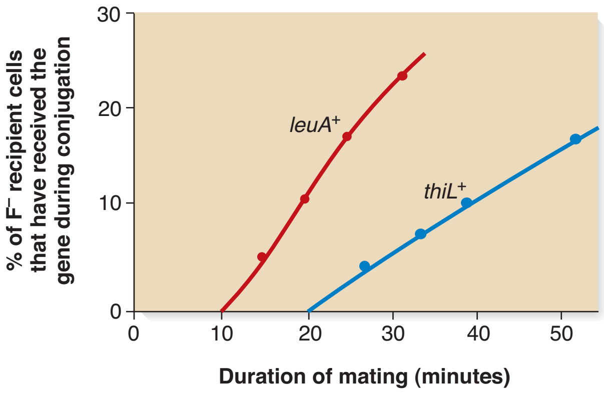
Answer:

A. In comparing the data among different *Hfr* strains, the order of the nine genes was always the same or the reverse of the same order. For example, *HfrH* and *Hfr4* transfer the same genes but their orders are reversed relative to each other. In addition, the *Hfr* strains showed an overlapping pattern of transfer with regard to the origin. For example, *Hfr1* and *Hfr2* had the same order of genes, but *Hfr1* began with *leu* and ended with *azi*, whereas *Hfr2* began with *pro* and ended with *lac*. From these findings, Wollman and Jacob concluded that the segment of DNA that was the origin of transfer had been inserted at different points within a circular *E. coli* chromosome in different *Hfr* strains. They also concluded that the origin can be inserted in either orientation, so the direction of gene transfer can be clockwise or counterclockwise around the circular bacterial chromosome.

B. A genetic map consistent with these results is shown here.



S3. An *Hfr* strain that is *leuA*$^+$ and *thiL*$^+$ was mated to a strain that is *leuA*$^-$ and *thiL*$^-$. In the data points shown here, the mating was interrupted, and the percentage of recombinants for each gene was determined by streaking on media that lacked either leucine or thiamine. The results are shown here.



What is the map distance (in minutes) between these two genes?

Answer: This problem is solved by extrapolating the data points to the *x*-axis to determine the time of entry. For *leuA*$^+$, they extrapolate back to 10 minutes. For *thiL*$^+$, they extrapolate back to 20 minutes. Therefore, the distance between the two genes is approximately 10 minutes.

S4. Genetic transfer via transformation can also be used to map genes along the bacterial chromosome. In this approach, fragments of chromosomal DNA are isolated from one bacterial strain and used to transform another strain. The experimenter examines the transformed bacteria to see if they have incorporated two or more different genes. For example, the DNA may be isolated from a donor *E. coli* bacterium that has functional copies of the *araB* and *leuD* genes. Let's call these genes *araB*$^+$ and *leuD*$^+$ to indicate the genes are functional. These two genes are required for arabinose metabolism and leucine synthesis, respectively. To map the distance between these two genes via transformation, a recipient bacterium is used that is *araB*$^-$ and *leuD*$^-$. Following transformation, the recipient bacterium may become *araB*$^+$ and *leuD*$^+$. This phenomenon is called cotransformation because two genes from the donor bacterium have been transferred to the recipient via transformation. In this type of experiment, the recipient cell is exposed to a fairly low concentration of donor DNA, making it unlikely that the recipient bacterium will take up more than one fragment of DNA. Therefore, under these conditions, cotransformation is likely only when two genes are fairly close together and are found on one fragment of DNA.

In a cotransformation experiment, a researcher has isolated DNA from an *araB*$^+$ and *leuD*$^+$ donor strain. This DNA was transformed into a recipient strain that was *araB*$^-$ and *leuD*$^-$. Following transformation, the cells were plated on a medium containing arabinose and leucine. On this medium, only bacteria that are *araB*$^+$ can grow. The bacteria can be either *leuD*$^+$ or *leuD*$^-$ because leucine is provided in the medium. Colonies that grew on this medium were then restreaked on a medium that contained arabinose but lacked leucine. Only *araB*$^+$ and *leuD*$^+$ cells could grow on these secondary plates. Following this protocol, a researcher obtained the following results:

Number of colonies growing on arabinose plus leucine medium: 57

Number of colonies that grew when restreaked on an arabinose medium without leucine: 42

What is the map distance between these two genes? *Note:* This problem can be solved using the strategy of a cotransduction experiment except that the researcher must determine the average size of DNA fragments that are taken up by the bacterial cells. This would correspond to the value of *L* in a cotransduction experiment.

Answer: As mentioned, the basic principle of gene mapping via cotransformation is identical to the method of gene mapping via cotransduction described in this chapter. One way to calculate the map distance is to use the same equation that we used for cotransduction data, except that we substitute cotransformation frequency for cotransduction frequency.

$$\text{Cotransformation frequency} = (1 - d/L)^3$$

(Note: Cotransformation is not quite as accurate as cotransduction because the sizes of chromosomal pieces tend to vary significantly from experiment to experiment, so the value of *L* is not quite as reliable. Nevertheless, cotransformation has been used extensively to map the order and distance between closely linked genes along the bacterial chromosome.)

The researcher needs to experimentally determine the value of *L* by running the DNA on a gel and estimating the average size of the DNA fragments. Let's assume they are about 2% of the bacterial chromosome, which, for *E. coli*, would be about 80,000 bp in length. So *L* equals 2 minutes, which is the same as 2%.

$$\text{Cotransformation frequency} = (1 - d/\text{L})^3$$

$$42/57 = (1 - d/2)^3$$

$$d = 0.2 \text{ minutes}$$

The distance between *araB* and *leuD* is approximately 0.2 minutes.

S5. In our discussion of transduction via P1 or P22, the reproductive cycle of the bacteriophage sometimes resulted in the packaging of a piece of the bacterial chromosome. For other bacteriophages, however, transduction may involve the transfer of only a few specific genes from the donor cell to the recipient. This phenomenon is known as specialized transduction. The key event that causes specialized transduction to occur is that the lysogenic phase of the phage reproductive cycle involves the integration of the viral DNA at a single specific site within the bacterial chromosome. The transduction of particular bacterial genes involves an abnormal excision of the phage DNA from this site within the chromosome that carries adjacent bacterial genes. For example, a bacteriophage called lambda (λ) that infects *E. coli* specifically integrates between two genes designated gal^+ and bio^+ (required for galactose utilization and biotin synthesis, respectively). Either of these genes could be packaged into the phage if an abnormal excision event occurred. How would specialized transduction be different from generalized transduction?

Answer: Generalized transduction can involve the transfer of any bacterial gene, but specialized transduction can transfer only genes that are adjacent to the site where the phage integrates. As mentioned, a bacteriophage that infects *E. coli* cells, known as lambda (λ), provides a well-studied example of specialized transduction. In the case of phage λ, the lysogenic reproductive cycle results in the integration of the phage DNA at a specific site that is located between two bacterial genes, gal^+ and bio^+. An *E. coli* strain that is lysogenic for phage λ has the lambda DNA integrated between these two bacterial genes. On occasion, the phage may enter the lytic cycle and excise its DNA from the bacterial chromosome. When this occurs normally, the phage excises its entire viral DNA from the bacterial chromosome. The excised phage DNA is then replicated and becomes packaged into newly made phages. However, an abnormal excision does occur at a low rate (i.e., about one in a million). In this abnormal event, the phage DNA is excised in such a way that an adjacent bacterial gene is included and some of the phage DNA is not included in the final product. For example, the abnormal excision may yield a fragment of DNA that includes the gal^+ gene and some of the lambda DNA but is missing part of the lambda DNA. If this DNA fragment is packaged into a virus, it is called a defective phage because it is missing some of the phage DNA. If it carries the gal^+ gene, it is designated λ*dgal* (the letter *d* designates a defective phage). Alternatively, an abnormal excision may carry the bio^+ gene. This phage is designated λ*dbio*. Defective λ phages can then transduce the gal^+ or bio^+ genes to other *E. coli* cells.

Conceptual Questions

C1. The terms *conjugation, transduction,* and *transformation* are used to describe three different natural forms of genetic transfer between bacterial cells. Briefly discuss the similarities and differences among these processes.

C2. Conjugation is sometimes called "bacterial mating." Is it a form of sexual reproduction? Explain.

C3. If you mix together an equal number of F^+ and F^- cells, how would you expect the proportions to change over time? In other words, do you expect an increase in the relative proportions of F^+ or of F^- cells? Explain your answer.

C4. What is the difference between an F^+ and an *Hfr* strain? Which type of strain do you expect to transfer many bacterial genes to recipient cells?

C5. What is the role of the origin of transfer during F^+- and Hfr-mediated conjugation? What is the significance of the direction of transfer in Hfr-mediated conjugation?

C6. What is the role of sex pili during conjugation?

C7. Think about the structure and transmission of F factors and discuss how you think F factors may have originated.

C8. What is cotransduction? What determines the likelihood that two genes will be cotransduced?

C9. When bacteriophage P1 causes *E. coli* to lyse, the resulting material is called a P1 lysate. What type of genetic material would be found in most of the P1 phages in the lysate? What kind of genetic material is occasionally found within a P1 phage?

C10. As described in Figure 9.10, host DNA is hydrolyzed into small pieces, which are occasionally assembled with phage proteins, creating a phage with bacterial chromosomal DNA. If the breakage of the chromosomal DNA is not random (i.e., it is more likely to break at certain spots as opposed to other spots), how might nonrandom breakage affect cotransduction frequency?

C11. Describe the steps that occur during bacterial transformation. What is a competent cell? What factors may determine whether a cell will be competent?

C12. Which bacterial genetic transfer process does not require recombination with the bacterial chromosome?

C13. Researchers who study the molecular mechanism of transformation have identified many proteins in bacteria that function in the uptake of DNA from the environment and its recombination into the host cell's chromosome. This means that bacteria have evolved molecular mechanisms for the purpose of transformation by extracellular DNA. Of what advantage(s) is it for a bacterium to import DNA from the environment and/or incorporate it into its chromosome?

C14. Antibiotics such as tetracycline, streptomycin, and bacitracin are small organic molecules that are synthesized by particular species of bacteria. Microbiologists have hypothesized that the reason why certain bacteria make antibiotics is to kill other species that occupy the same environment. Bacteria that produce an antibiotic may be able to kill competing species. This provides more resources for the antibiotic-producing bacteria. In addition, bacteria that have the genes necessary for antibiotic biosynthesis contain genes that confer resistance to the same antibiotic. For example, tetracycline is made by the soil bacterium *Streptomyces aureofaciens*. Besides the genes that are needed to make tetracycline, *S. aureofaciens* also contains genes that confer tetracycline resistance; otherwise, it would kill itself when it makes tetracycline. In recent years, however, many other species of bacteria that do not synthesize tetracycline have acquired the genes that confer tetracycline resistance. For example, certain strains of *E. coli* carry tetracycline-resistance genes, even though *E. coli* does not synthesize tetracycline. When these genes were analyzed at the molecular level, it was found that they are evolutionarily related to the genes in *S. aureofaciens*. This observation indicates that the genes from *S. aureofaciens* have been transferred to *E. coli*.

A. What form of genetic transfer (i.e., conjugation, transduction, or transformation) would be the most likely mechanism of interspecies gene transfer?

B. Because *S. aureofaciens* is a nonpathogenic soil bacterium and *E. coli* is an enteric bacterium, do you think it was direct gene transfer, or do you think it may have occurred in multiple steps (i.e., from *S. aureofaciens* to other bacterial species and then to *E. coli*)?

C. How could the widespread use of antibiotics to treat diseases have contributed to the proliferation of many bacterial species that are resistant to antibiotics?

Application and Experimental Questions

E1. In the experiment of Figure 9.1, a met^- bio^- thr^+ leu^+ thi^+ cell could become met^+ bio^+ thr^+ leu^+ thi^+ by a (rare) double mutation that converts the met^- bio^- genetic material into met^+ bio^+. Likewise, a met^+ bio^+ thr^- leu^- thi^- cell could become met^+ bio^+ thr^+ leu^+ thi^+ by three mutations that convert the thr^- leu^- thi^- genetic material into thr^+ leu^+ thi^+. From the results of Figure 9.1, how do you know that the occurrence of 10 met^+ bio^+ thr^+ leu^+ thi^+ colonies is not due to these types of rare double or triple mutations?

E2. In the experiment of Figure 9.1, Joshua Lederberg and Edward Tatum could not discern whether met^+ bio^+ genetic material was transferred to the met^- bio^- thr^+ leu^+ thi^+ strain or if thr^+ leu^+ thi^+ genetic material was transferred to the met^+ bio^+ thr^- leu^- thi^- strain. Let's suppose that one strain is streptomycin-resistant (say, met^+ bio^+ thr^- leu^- thi^-) and the other strain is sensitive to streptomycin. Describe an experiment that could determine whether the met^+ bio^+ genetic material was transferred to the met^- bio^- thr^+ leu^+ thi^+ strain or if the thr^+ leu^+ thi^+ genetic material was transferred to the met^+ bio^+ thr^- leu^- thi^- strain.

E3. Explain how a U-tube apparatus can distinguish between genetic transfer involving conjugation and genetic transfer involving transduction. Do you think a U-tube could be used to distinguish between transduction and transformation?

E4. What is an interrupted mating experiment? What type of experimental information can be obtained from this type of study? Why is it necessary to interrupt mating?

E5. In a conjugation experiment, what is meant by the time of entry? How is the time of entry determined experimentally?

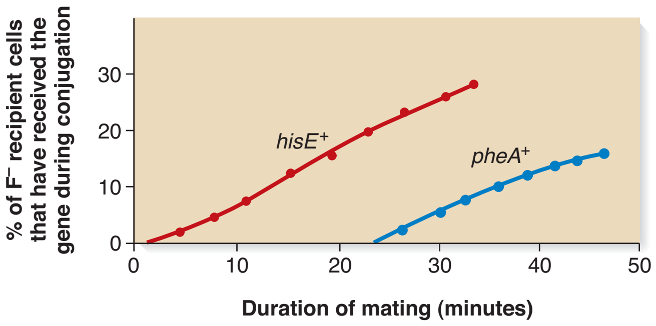
E6. In your laboratory, you have an F^- strain of *E. coli* that is resistant to streptomycin and is unable to metabolize lactose, but it can metabolize glucose. Therefore, this strain can grow on media that contain glucose and streptomycin, but it cannot grow on media containing lactose. A researcher has sent you two *E. coli* strains in two separate tubes. One strain, let's call it strain *A*, has an F factor that carries the genes that are required for lactose metabolism. On its chromosome, it also has the genes that are required for glucose metabolism. However, it is sensitive to streptomycin. This strain can grow on media containing lactose or glucose, but it cannot grow if streptomycin is added to the media. The second strain, let's call it strain *B*, is an F^- strain. On its chromosome, it has the genes that are required for lactose and glucose metabolism. Strain *B* is also sensitive to streptomycin. Unfortunately, when strains *A* and *B* were sent to you, the labels had fallen off the tubes. Describe how you could determine which tubes contain strain *A* and strain *B*.

E7. As mentioned in Solved Problem S2, origins of transfer can be located in many different locations, and their direction of transfer can be clockwise or counterclockwise. Let's suppose a researcher mated six different *Hfr* strains that were thr^+ leu^+ ton^s str^r azi^s lac^+ gal^+ pro^+ met^+ to an F^- strain that was thr^- leu^- ton^r str^s azi^r lac^- gal^- pro^- met^-, and obtained the following results:

Strain	Order of Gene Transfer
1	ton^s azi^s leu^+ thr^+ met^+ str^r gal^+ lac^+ pro^+
2	leu^+ azi^s ton^s pro^+ lac^+ gal^+ str^r met^+ thr^+
3	lac^+ gal^+ str^r met^+ thr^+ leu^+ azi^s ton^s pro^+
4	leu^+ thr^+ met^+ str^r gal^+ lac^+ pro^+ ton^s azi^s
5	ton^s pro^+ lac^+ gal^+ str^r met^+ thr^+ leu^+ azi^s
6	met^+ str^r gal^+ lac^+ pro^+ ton^s azi^s leu^+ thr^+

Draw a circular map of the *E. coli* chromosome and describe the locations and orientations of the origins of transfer in these six *Hfr* strains.

E8. An *Hfr* strain that is $hisE^+$ and $pheA^+$ was mated to a strain that is $hisE^-$ and $pheA^-$. The mating was interrupted and the percentage of recombinants for each gene was determined by streaking on media that lacked either histidine or phenylalanine. The following results were obtained:



A. Determine the map distance (in minutes) between these two genes.

B. In a previous experiment, it was found that *hisE* is 4 minutes away from the gene *pabB*. *PheA* was shown to be 17 minutes from this gene. Draw a genetic map describing the locations of all three genes.

E9. Acridine orange is a chemical that inhibits the replication of F factor DNA but does not affect the replication of chromosomal DNA, even if the chromosomal DNA contains an Hfr. Let's suppose that you have an *E. coli* strain that is unable to metabolize lactose and has an F factor that carries a streptomycin-resistant gene. You also have an F^- strain of *E. coli* that is sensitive to streptomycin and has the genes that allow the bacterium to metabolize lactose. This second strain can grow on lactose-containing media. How would you generate an *Hfr* strain that is resistant to streptomycin and can metabolize lactose? (Hint: F factors occasionally integrate into the chromosome to become *Hfr* strains, and occasionally *Hfr* strains excise their DNA from the chromosome to become F^+ strains that carry an F′ factor.)

E10. In a P1 transduction experiment, the P1 lysate contains phages that carry pieces of the host chromosomal DNA, but the lysate also contains broken pieces of chromosomal DNA (see Figure 9.10). If a P1 lysate is used to transfer chromosomal DNA to another bacterium, how could you show experimentally that the recombinant bacterium has been transduced (i.e., taken up a P1 phage with a piece of chromosomal DNA inside) versus transformed (i.e., taken up a piece of chromosomal DNA that is not within a P1 phage coat)?

E11. Can you devise an experimental strategy to get P1 phage to transduce the entire lambda genome from one strain of bacterium to another strain? (*Note:* The general features of bacteriophage reproductive cycles are described in Chapter 10.) Phage λ has a genome size of 48,502 nucleotides (about 1% of the size of the *E. coli* chromosome) and can follow the lytic or lysogenic reproductive cycle. Growth of *E. coli* on minimal growth medium favors the lysogenic reproductive cycle, whereas growth on rich media and/or under UV light promotes the lytic cycle.

E12. Let's suppose a new strain of P1 has been identified that packages larger pieces of the *E. coli* chromosome. This new P1 strain packages pieces of the *E. coli* chromosome that are 5 minutes long. If two genes are 0.7 minutes apart along the *E. coli* chromosome, what would be the cotransduction frequency using a normal strain of P1 and using this new strain of P1 that packages large pieces? What would be the experimental advantage of using this new P1 strain?

E13. If two bacterial genes are 0.6 minutes apart on the bacterial chromosome, what frequency of cotransductants would you expect to observe in a P1 transduction experiment?

E14. In an experiment involving P1 transduction, the cotransduction frequency was 0.53. How far apart are the two genes?

E15. In a cotransduction experiment, the transfer of one gene is selected for and the presence of the second gene is then determined. If 0 out of 1000 P1 transductants that carry the first gene also carry the second gene, what would you conclude about the minimum distance between the two genes?

E16. In a cotransformation experiment (see Solved Problem S4), DNA was isolated from a donor strain that was $proA^+$ and $strC^+$ and sensitive to tetracycline. (The *proA* and *strC* genes confer the ability to synthesize proline and confer streptomycin resistance, respectively.) A recipient strain is $proA^-$ and $strC^-$ and is resistant to tetracycline. After transformation, the bacteria were first streaked on a medium containing proline, streptomycin, and tetracycline. Colonies were then restreaked on a medium containing streptomycin and tetracycline. (Note: Both types of media had carbon and nitrogen sources for growth.) The following results were obtained:

> 70 colonies grew on the medium containing proline, streptomycin, and tetracycline, but only 2 of these 70 colonies grew when restreaked on the medium containing streptomycin and tetracycline but lacking proline.

A. If we assume the average size of the DNA fragments is 2 minutes, how far apart are these two genes?

B. What would you expect the cotransformation frequency to be if the average size of the DNA fragments was 4 minutes and the two genes are 1.4 minutes apart?

Questions for Student Discussion/Collaboration

1. Discuss the advantages of the genetic analysis of bacteria. Make a list of the types of allelic differences among bacteria that are suitable for genetic analyses.
2. Of the three types of genetic transfer, discuss which one(s) is/are more likely to occur between members of different species. Discuss some of the potential consequences of interspecies genetic transfer.

Answers to Comprehension Questions

9.1: c

9.2: a, d

9.3: c, b

9.4: a, b

9.5: c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

10

CHAPTER OUTLINE

10.1 Virus Structure and Genetic Composition
10.2 Viral Reproductive Cycles
10.3 Plaque Formation and Intergenic Complementation in Bacteriophages
10.4 Intragenic Mapping in Bacteriophages

A plant infected with tobacco mosaic virus.

GENETICS OF VIRUSES

Viruses are nonliving particles with nucleic acid genomes. Why are viruses considered nonliving? They do not exhibit all of the properties associated with living organisms. For example, viruses are not composed of cells, and by themselves, they do not carry out metabolism, use energy, maintain homeostasis, or even reproduce. A virus or its genetic material must be taken up by a living cell to replicate.

The first virus to be discovered was tobacco mosaic virus (TMV). This virus infects many species of plants and causes mosaic-like patterns in which normal-colored patches are interspersed with light green or yellowish patches on the leaves (see chapter opening photo). TMV damages leaves, flowers, and fruit, but almost never kills the plant. In 1883, the German scientist Adolf Mayer determined that this disease could be spread by spraying the sap from one plant onto another. By subjecting this sap to filtration, the Russian scientist Dmitri Ivanovski demonstrated that the disease-causing agent was not a bacterium. Sap that had been passed through filters with pores small enough to prevent the passage of bacterial cells was still able to spread the disease. At first, some researchers suggested the agent was a chemical toxin. However, the Dutch botanist Martinus Beijerinck ruled out this possibility by showing that sap could continue to transmit the disease after many plant generations. A toxin would have been diluted after many generations, whereas Beijerinck's results indicated the disease agent was multiplying in the plant. Around the same time, animal viruses were discovered in connection with a livestock infection called foot-and-mouth disease. In 1900, the first human virus, the virus that causes yellow fever, was identified.

Since these early studies, microbiologists, geneticists, and molecular biologists have taken a great interest in the structure, genetic composition, and replication of viruses. All organisms are susceptible to infection by viruses. Once a cell is infected, the genetic material of a virus orchestrates a series of events that ultimately leads to the production of new virus particles. In this chapter, we will first examine the structure and genetic composition of viruses and explore viral reproductive cycles. We will then examine some pivotal studies of a virus called T4, which infects the bacterium *Escherichia coli*. Studies of this virus helped researchers gain a better understanding of gene structure.

10.1 VIRUS STRUCTURE AND GENETIC COMPOSITION

A **virus** is a small infectious particle that consists of one or more nucleic acid molecules enclosed in a protein coat. Researchers have identified and studied over 4000 different types of viruses. Although all viruses share some similarities, such as small size and the reliance on a living cell for replication, they vary greatly in their characteristics, including their host range, structure, and genome composition. Some of the major differences are described next, and characteristics of selected viruses are shown in **Table 10.1**.

Differences in Host Range A cell that is infected by a virus is called a **host cell,** and a species that can be infected by a specific virus is called a host species for that virus. Viruses differ greatly in their **host range**—the number of species and cell types they can infect. Table 10.1 lists a few examples of viruses with widely different ranges of host species. Tobacco mosaic virus, which we discussed earlier, has a broad host range. TMV is known to infect over 150 different species of plants. By comparison, other viruses have a narrow host range, with some infecting only a single species. Furthermore, a virus may infect only certain cell types in a host species. For example, herpesviruses specifically infect nerve cells.

Structural Differences Although the existence of viruses was postulated in the 1890s, viruses were not observed until the 1930s, when the electron microscope was invented. Viruses cannot be resolved by even the best light microscope. Most of them are smaller than the wavelength of visible light. Viruses range in size from about 20–400 nm in diameter (1 nanometer = 10^{-9} meters). For comparison, a typical bacterium is 1000 nm in diameter, and the diameter of most eukaryotic cells is 10 to 1000 times that of a bacterium. Adenoviruses, which cause infections of the respiratory and gastrointestinal tracts, have an average diameter of 75 nm. Over 50 million adenoviruses could fit into an average-sized human cell.

What are the common structural features of all viruses? As shown in **Figure 10.1**, all viruses have a protein coat called a **capsid** that encloses a genome consisting of one or more molecules of nucleic acid. Capsids are composed of one or several different protein subunits called **capsomers.** Capsids have a variety of shapes, including helical and polyhedral. Figure 10.1a shows the structure of TMV, which has a helical capsid made of identical capsomers. Figure 10.1b shows an adenovirus, which has a polyhedral capsid. Protein fibers with a terminal knob are located at the corners of the polyhedral capsid. Many viruses that infect animal cells, such as the influenza virus shown in Figure 10.1c, have a **viral envelope** enclosing the capsid. The envelope consists of a lipid bilayer that is derived from the plasma membrane of the host cell and is embedded with virally encoded spike glycoproteins, also called spikes or peplomers. In addition to encasing and protecting the genetic material, the capsid and envelope enable viruses to infect their hosts. In many viruses, the capsids or envelopes have specialized proteins, including protein fibers with a knob (Figure 10.1b) and spike glycoproteins (Figure

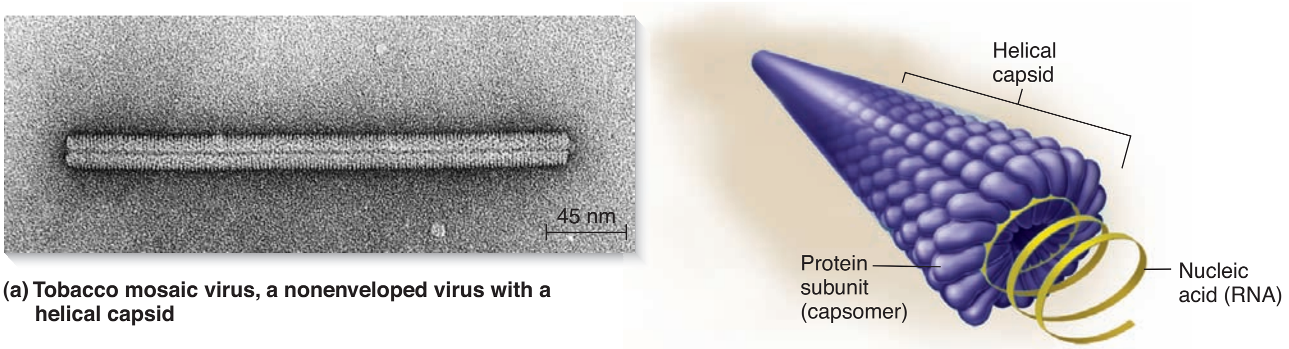
TABLE **10.1**

Hosts and Characteristics of Selected Viruses

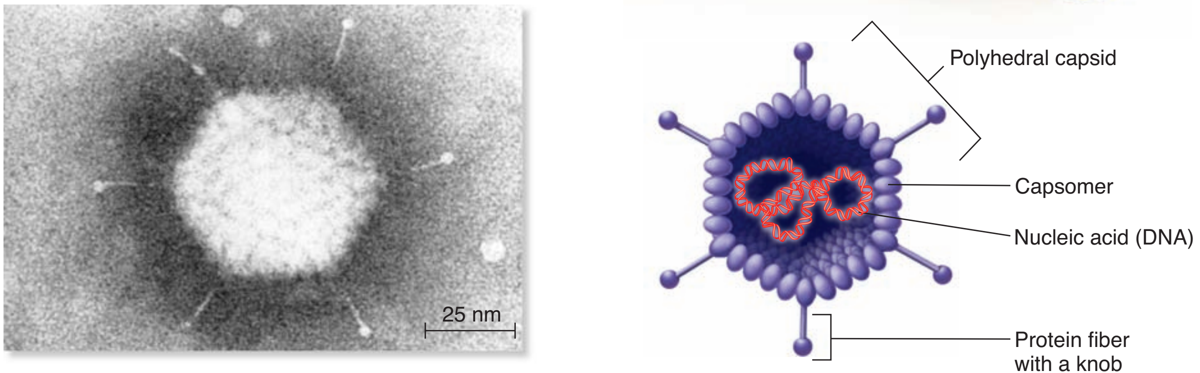
Virus or Group of Viruses	Host	Effect on Host	Nucleic Acid*	Genome Size (kb)[†]	Number of Genes[†]
Phage fd	*E. coli*	Slows growth	ssDNA	6.4	10
Phage λ	*E. coli*	Can exist harmlessly in the host cell or cause lysis	dsDNA	48.5	71
Phage T4	*E. coli*	Causes lysis	dsDNA	169	288
Phage Qß	*E. coli*	Slows growth	ssRNA	4.2	4
Tobacco mosaic virus (TMV)	Many plants	Causes mottling and necrosis of leaves and other plant parts	ssRNA	6.4	6
Baculoviruses	Insects	Most baculoviruses are species-specific; they usually kill the insect.	dsDNA	133.9	154
Parvovirus	Mammals	Causes respiratory, flulike symptoms	ssDNA	5.0	5
Influenza virus	Mammals	Causes classic flu symptoms—fever, cough, sore throat, and headache	ssRNA	13.5	11
Epstein-Barr virus	Humans	Causes mononucleosis, with fever, sore throat, and fatigue	dsDNA	172	80
Adenovirus	Humans	Causes respiratory symptoms and diarrhea	dsDNA	34	35
Herpes simplex type II	Humans	Causes blistering sores around the genital region	dsDNA	158.4	77
HIV	Humans	Causes AIDS, an immunodeficiency syndrome eventually leading to death	ssRNA	9.7	9

*The abbreviations ss and ds refer to single-stranded and double-stranded, respectively.

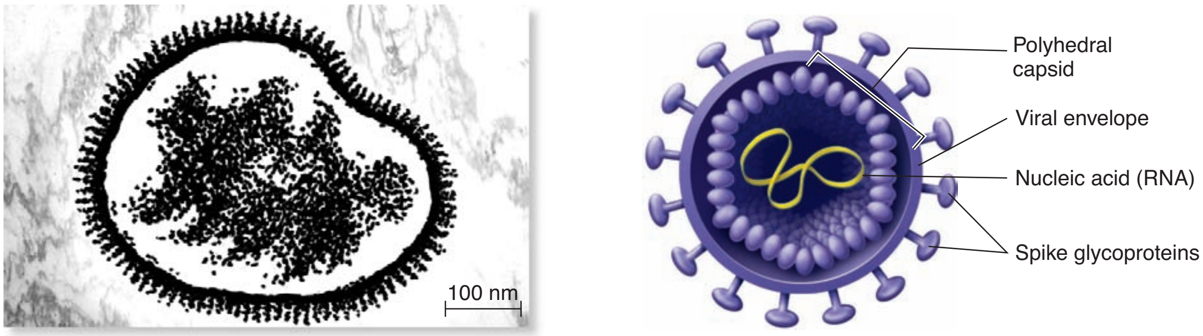
[†] Several of the viruses listed in this table are found in different strains that show variation with regard to genome size and number of genes. The numbers reported in this table are typical values. The abbreviation kb refers to kilobase, which equals 1000 bases.


45 nm
Helical capsid
Protein subunit (capsomer)
Nucleic acid (RNA)

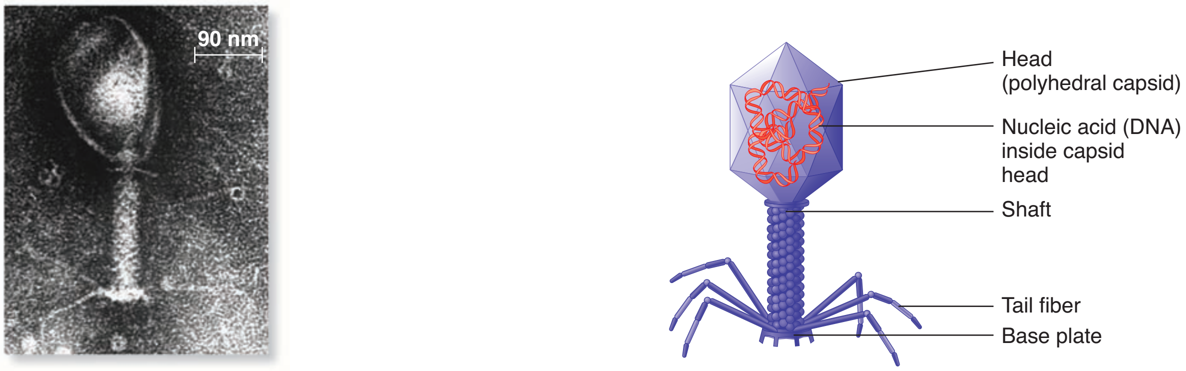
(a) Tobacco mosaic virus, a nonenveloped virus with a helical capsid


25 nm
Polyhedral capsid
Capsomer
Nucleic acid (DNA)
Protein fiber with a knob

(b) Adenovirus, a nonenveloped virus with a polyhedral capsid and protein fibers with a knob


100 nm
Polyhedral capsid
Viral envelope
Nucleic acid (RNA)
Spike glycoproteins

(c) Influenza virus, an enveloped virus with spikes


90 nm
Head (polyhedral capsid)
Nucleic acid (DNA) inside capsid head
Shaft
Tail fiber
Base plate

(d) T4, a bacteriophage

FIGURE 10.1 **Variations in the structure of viruses, as shown by transmission electron microscopy.** All viruses contain nucleic acid (DNA or RNA) surrounded by a protein capsid. They may or may not have an outer envelope surrounding the capsid. **(a)** Tobacco mosaic virus (TMV) has a capsid made of 2130 identical protein subunits, helically arranged around a strand of RNA. **(b)** Adenoviruses have polyhedral capsids containing protein fibers with a knob. **(c)** Many animal viruses, including the influenza virus, have an envelope composed of a lipid bilayer and spike glycoproteins. The lipid bilayer is obtained from the host cell when the virus buds from the plasma membrane. **(d)** Some bacteriophages, such as T4, have capsids with accessory structures, such as tail fibers and base plates, that facilitate invasion of a bacterial cell.

Concept check: What features vary among different types of viruses?

10.1c), that help them bind to the surface of a host cell. Viruses that infect bacteria, called **bacteriophages,** or **phages,** may have more complex capsids, with accessory structures used for anchoring the virus to a host cell and injecting the viral nucleic acid (Figure 10.1d). As discussed later, the tail fibers of such bacteriophages attach the virus to the bacterial cell wall.

Genome Differences The genetic material in a virus is called a **viral genome.** The composition of viral genomes varies markedly among different types of viruses, as suggested by the examples in Table 10.1. The nucleic acid of some viruses is DNA, whereas in others it is RNA. These are referred to as DNA viruses and RNA viruses, respectively. It is striking that some viruses use RNA for their genome, whereas all living organisms use DNA. In some viruses, the nucleic acid is single-stranded, whereas in others, it is double-stranded. The genome can be linear or circular, depending on the type of virus. Some kinds of viruses have more than one copy of the genome. Viral genomes also vary considerably in size, ranging from a few thousand to more than a hundred thousand nucleotides in length (see Table 10.1). For example, the genomes of some simple viruses, such as phage Qβ, are only a few thousand nucleotides in length and contain only a few genes. Other viruses, particularly those with a complex structure, such as phage T4, contain many more genes. These extra genes encode many different proteins that are involved in the formation of the elaborate structure shown in Figure 10.1d.

10.1 REVIEWING THE KEY CONCEPTS

- Tobacco mosaic virus (TMV) was the first virus to be discovered. It infects many species of plants.
- Viruses vary with regard to their host range, structure, and genome composition (see Table 10.1, Figure 10.1).

10.1 COMPREHENSION QUESTIONS

1. What are common features found in all viruses?
   a. An envelope
   b. DNA
   c. Nucleic acid surrounded by a protein capsid
   d. All of the above
2. Viral genomes can be
   a. DNA or RNA.
   b. single-stranded or double-stranded.
   c. linear or circular.
   d. all of the above.

10.2 VIRAL REPRODUCTIVE CYCLES

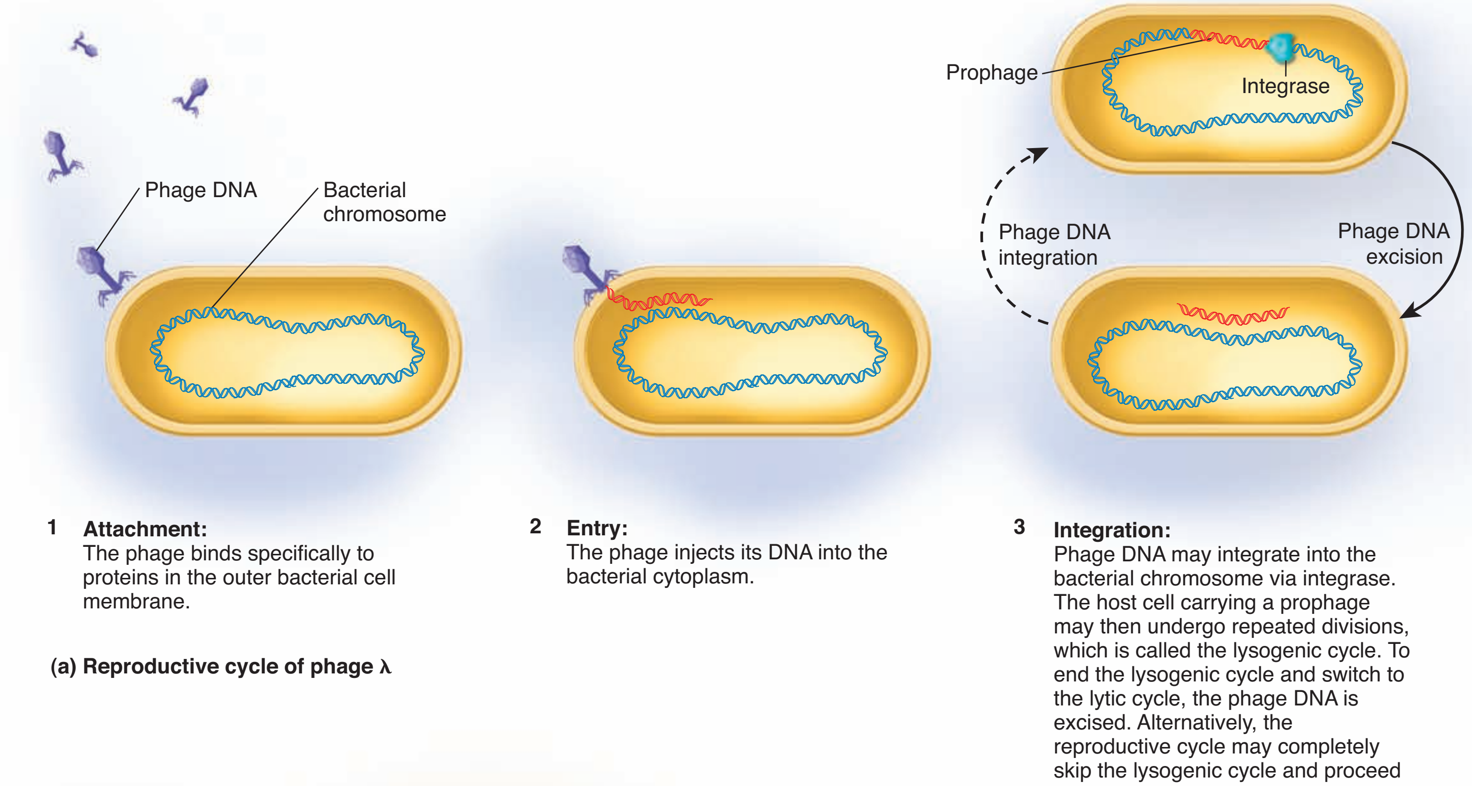
When a virus infects a host cell, the expression of viral genes leads to a series of steps, called a **viral reproductive cycle** that results in the production of new viruses. The details of the steps may be quite different among various types of viruses, and even the same virus may have the capacity to follow alternative cycles. Even so, by studying the reproductive cycles of hundreds of different viruses, researchers have determined that the viral reproductive cycle consists of five or six basic steps. To illustrate the general features of viral reproductive cycles, **Figure 10.2** considers these steps for two types of viruses. Figure 10.2a shows the cycle of **phage λ** (lambda), a bacteriophage with double-stranded DNA as its genome, and Figure 10.2b depicts the cycle of **human immunodeficiency virus (HIV),** an enveloped animal virus containing single-stranded RNA. The descriptions that follow compare the reproductive cycles of these two very different viruses.

Step 1: Attachment In the first step of a viral reproductive cycle, the virus must attach to the surface of a host cell. This attachment is usually specific for one or just a few types of cells because proteins in the virus recognize and bind to specific molecules on the cell surface. In the case of phage λ, the phage tail fibers bind to proteins in the outer bacterial cell membrane of *E. coli* cells. In the case of HIV, spike glycoproteins in the viral envelope bind to receptors in the plasma membrane of human blood cells called helper T cells.

Step 2: Entry After attachment, the viral genome enters the host cell. Attachment of phage λ stimulates a conformational change in the phage coat proteins, so the shaft (also called the sheath) contracts, and the phage injects its DNA into the bacterial cytoplasm. In contrast, the envelope of HIV fuses with the plasma membrane of the host cell, so both the capsid and its contents are released into the cytosol. Some of the HIV capsid proteins are then removed by host cell enzymes, a process called uncoating. This releases two copies of the viral RNA and two molecules of an enzyme called reverse transcriptase into the cytosol. Reverse transcriptase will be needed for step 3. Once a viral genome has entered the cell, specific viral genes are expressed immediately due to the action of host cell enzymes and ribosomes. Expression of these key genes leads quickly to either step 3 or step 4 of the reproductive cycle, depending on the type of virus. The genome of some viruses, including both phage λ and HIV, can integrate into a chromosome of the host cell. For such viruses, the cycle may proceed from step 2 to step 3 as described next, delaying the production of new viruses. Alternatively, the cycle may proceed directly from step 2 to step 4 and quickly lead to the production of new viruses.

Step 3: Integration Viruses capable of integration carry a gene that encodes an enzyme called **integrase.** For integration to occur, this gene is expressed soon after entry so that the integrase protein is made. Integrase cuts the host's chromosomal DNA and inserts the viral genome into the chromosome. In the case of phage λ, the double-stranded DNA that entered the cell can be directly integrated into the double-stranded DNA of the chromosome. Once integrated, the phage DNA in a bacterium is called a **prophage.** While it exists as a prophage, this type of viral reproductive cycle is called the **lysogenic cycle.** As discussed later, new phages are not made during the lysogenic cycle, and the host cell is not destroyed. On occasion, a prophage can be excised from the bacterial chromosome and proceed to step 4.

How can an RNA virus integrate its genome into the host cell's DNA? For this to occur, the viral genome must be copied into DNA. HIV accomplishes this by means of a viral enzyme

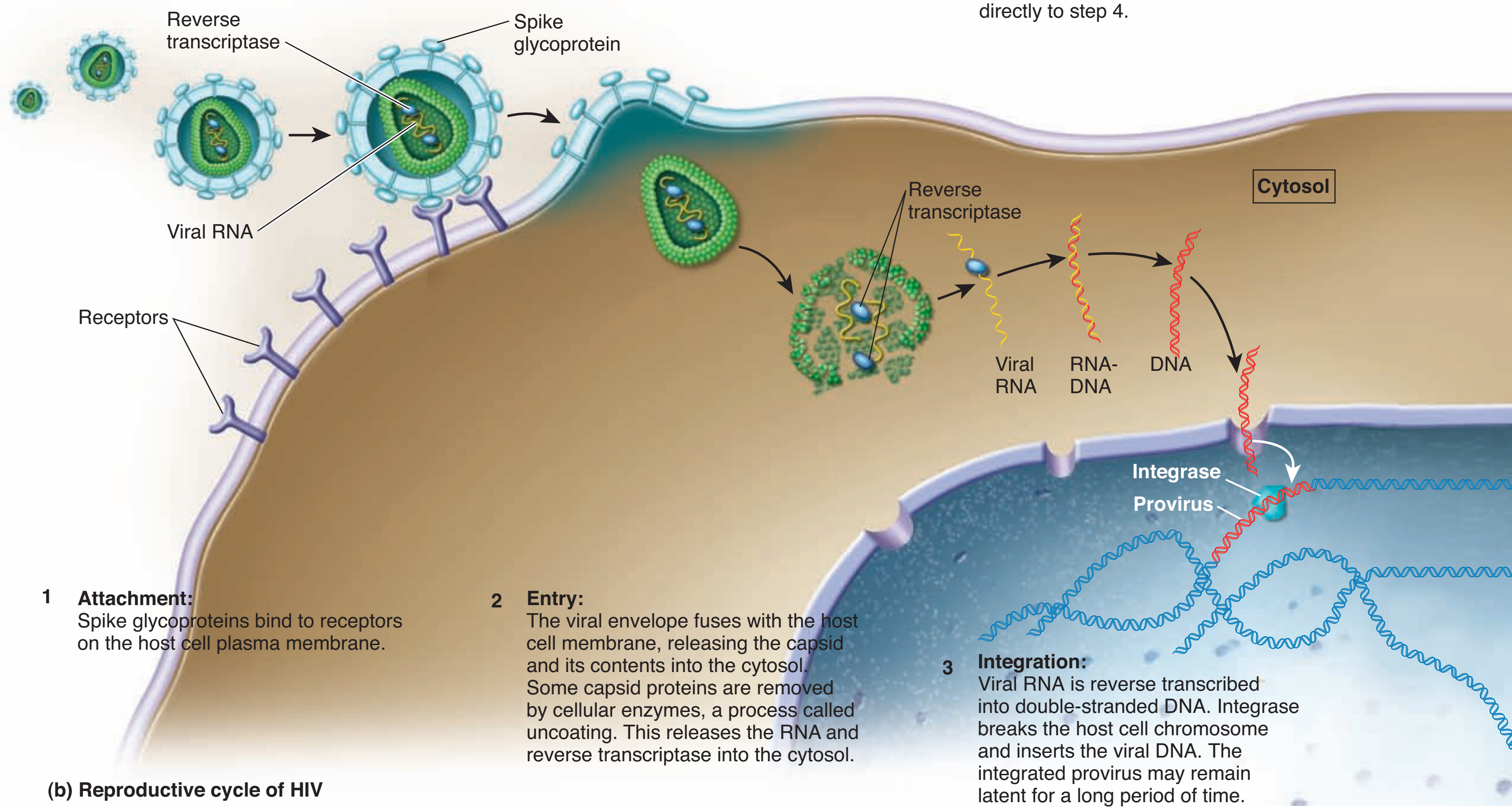

Phage DNA
Bacterial chromosome
Prophage
Integrase
Phage DNA integration
Phage DNA excision

1 **Attachment:**
The phage binds specifically to proteins in the outer bacterial cell membrane.

2 **Entry:**
The phage injects its DNA into the bacterial cytoplasm.

3 **Integration:**
Phage DNA may integrate into the bacterial chromosome via integrase. The host cell carrying a prophage may then undergo repeated divisions, which is called the lysogenic cycle. To end the lysogenic cycle and switch to the lytic cycle, the phage DNA is excised. Alternatively, the reproductive cycle may completely skip the lysogenic cycle and proceed directly to step 4.

(a) Reproductive cycle of phage λ


Reverse transcriptase
Spike glycoprotein
Viral RNA
Receptors
Reverse transcriptase
Cytosol
Viral RNA
RNA-DNA
DNA
Integrase
Provirus

1 **Attachment:**
Spike glycoproteins bind to receptors on the host cell plasma membrane.

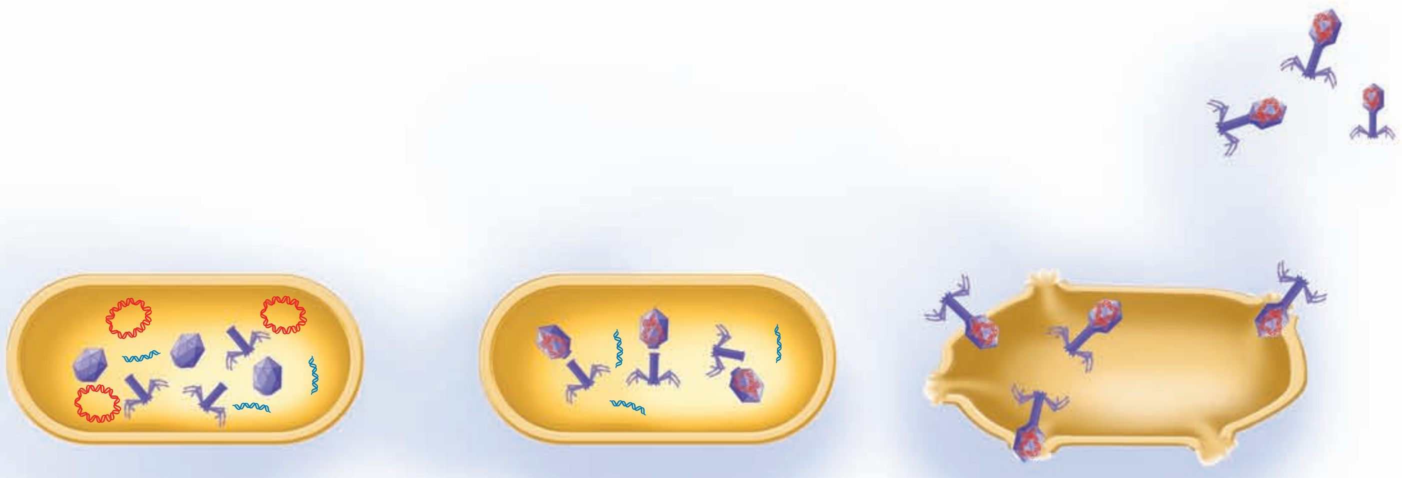
2 **Entry:**
The viral envelope fuses with the host cell membrane, releasing the capsid and its contents into the cytosol. Some capsid proteins are removed by cellular enzymes, a process called uncoating. This releases the RNA and reverse transcriptase into the cytosol.

3 **Integration:**
Viral RNA is reverse transcribed into double-stranded DNA. Integrase breaks the host cell chromosome and inserts the viral DNA. The integrated provirus may remain latent for a long period of time.

(b) Reproductive cycle of HIV

FIGURE 10.2 **Comparison of the steps of two viral reproductive cycles.** (**a**) The reproductive cycle of phage λ, a bacteriophage with a double-stranded DNA genome. (**b**) The reproductive cycle of HIV, an enveloped animal virus with a single-stranded RNA genome.

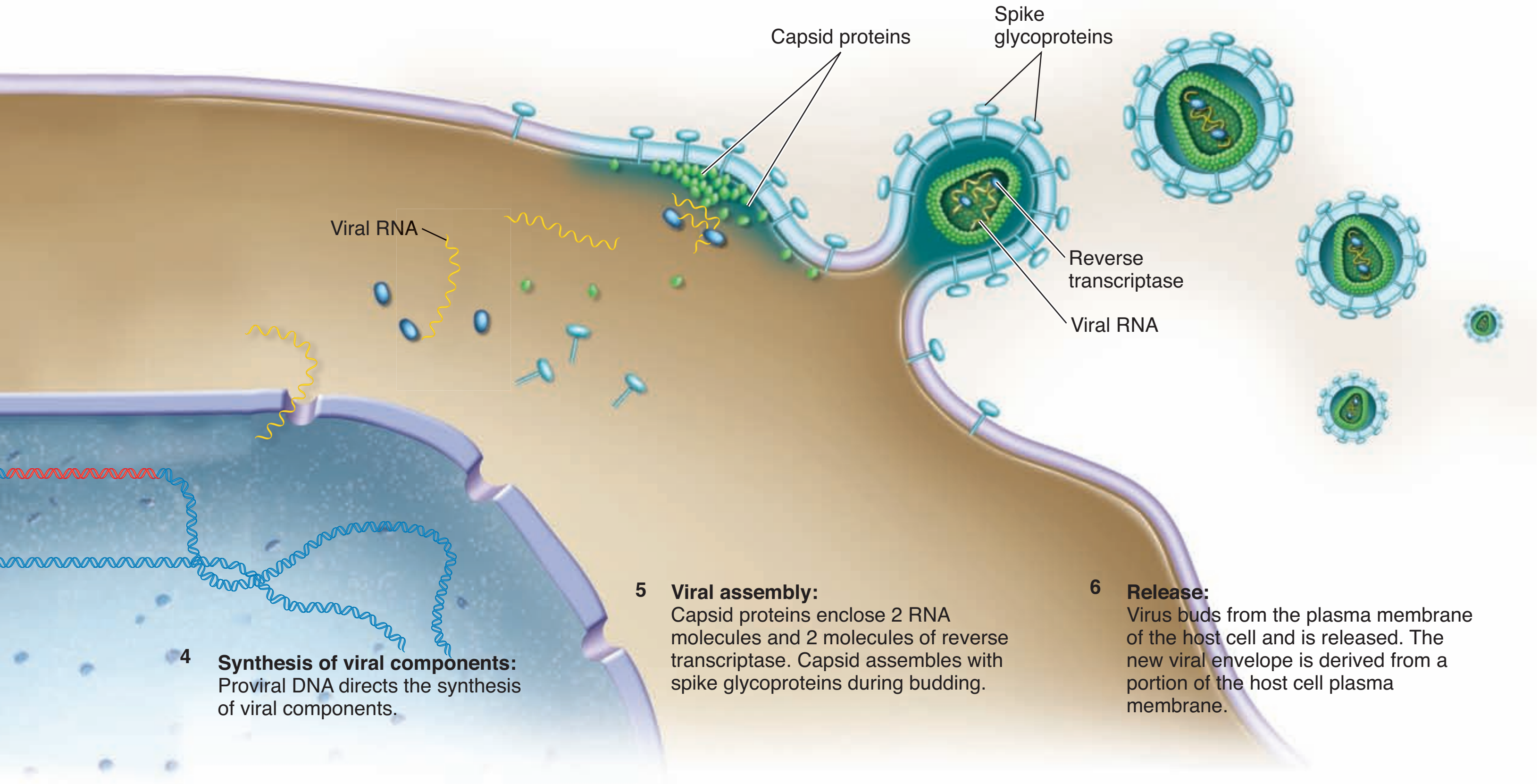
Concept check: *During which step may a virus remain latent?*



4 **Synthesis of viral components:** In the lytic cycle, phage DNA directs the synthesis of viral components. During this process, the phage DNA circularizes, and the host chromosomal DNA is degraded.

5 **Viral assembly:** Phage components are assembled with the help of noncapsid proteins to make many new phages.

6 **Release:** The viral enzyme called lysozyme causes cell lysis, and new phages are released from the broken cell.



4 **Synthesis of viral components:** Proviral DNA directs the synthesis of viral components.

5 **Viral assembly:** Capsid proteins enclose 2 RNA molecules and 2 molecules of reverse transcriptase. Capsid assembles with spike glycoproteins during budding.

6 **Release:** Virus buds from the plasma membrane of the host cell and is released. The new viral envelope is derived from a portion of the host cell plasma membrane.

called **reverse transcriptase,** which is carried within the capsid and released into the host cell along with the viral RNA. Reverse transcriptase uses the viral RNA strand to make a complementary copy of DNA, and it then uses the DNA strand as a template to make double-stranded viral DNA. This process is called reverse transcription because it is the reverse of the usual transcription process, in which a DNA strand is used to make a complementary strand of RNA. The viral double-stranded DNA enters the host cell nucleus and is inserted into a host chromosome via integrase. Once integrated, the viral DNA in a eukaryotic cell is called a **provirus.** Viruses that follow this mechanism are called retroviruses.

Step 4: Synthesis of Viral Components The production of new viruses by a host cell involves the replication of the viral genome and the synthesis of viral proteins that make up the capsid. In the case of a bacteriophage that has been integrated into the host chromosome, the prophage must be excised as described in step 3 before synthesis of new viral components can occur. A viral enzyme called excisionase is required for this process. Following excision, host cell enzymes make many copies of the phage DNA and transcribe the genes within these copies into mRNA. Host cell ribosomes translate this viral mRNA into viral proteins. The expression of phage genes also leads to the degradation of the host chromosomal DNA. In the case of HIV, the DNA provirus is not excised from the host chromosome. Instead, it is transcribed in the nucleus to produce many copies of viral RNA. These viral RNA molecules enter the cytosol, where they are used to make viral proteins and serve as the genome for new viral particles.

Step 5: Viral Assembly After all of the necessary components have been synthesized, they must be assembled into new viruses. Some viruses with a simple structure self-assemble, meaning that viral components spontaneously bind to each other to form a complete virus particle. An example of a self-assembling virus is TMV, which we examined earlier (see Figure 10.1a). The capsid proteins assemble around the RNA genome, which becomes trapped inside the hollow capsid (see Solved Problem S1 at the end of the chapter). This assembly process can occur in vitro if purified capsid proteins and RNA are mixed together. Other viruses, including the two shown in Figure 10.2, do not self-assemble. The correct assembly of phage λ requires the help of noncapsid proteins not found in the completed phage particle. Some of these noncapsid proteins function as enzymes that modify capsid proteins, and others serve as scaffolding for the assembly of the capsid. The assembly of an HIV virus occurs in two stages. First, capsid proteins assemble around two molecules of viral RNA and two molecules of reverse transcriptase. Next, the newly formed capsid acquires its outer envelope in a budding process. This second phase of assembly occurs during step 6, as the virus is released from the cell.

Step 6: Release The last step of a viral reproductive cycle is the release of new viruses from the host cell. The release of bacteriophages is a dramatic event. Because bacteria are surrounded by a rigid cell wall, the phages must burst, or lyse, their host cell in order to escape. After the phages have been assembled, a phage-encoded enzyme called lysozyme digests the bacterial cell wall, causing the cell to burst. Lysis releases many new phages into the environment, where they can infect other bacteria and begin the cycle again. Collectively, steps 1, 2, 4, 5, and 6 are called the **lytic cycle** because they lead to cell lysis.

The release of enveloped viruses from an animal cell is far less dramatic. This type of virus escapes by a mechanism called budding that does not lyse the cell. In the case of HIV, a newly assembled virus particle associates with a portion of the plasma membrane containing HIV spike glycoproteins. The membrane enfolds the viral capsid and eventually buds from the surface of the cell. This is how the virus acquires its envelope, which is a piece of host cell membrane studded with viral glycoproteins.

Latency in Bacteriophages As we saw in step 3, viruses can integrate their genomes into a host chromosome. In some cases, the prophage or provirus may remain inactive, or **latent,** for a long time. Most of the viral genes are silent during latency, and the viral reproductive cycle does not progress to step 4. Latency in bacteriophages is also called **lysogeny.** When this occurs, both the prophage and its host cell are said to be lysogenic. When a lysogenic bacterium prepares to divide, it copies the prophage DNA along with its own DNA, so each daughter cell inherits a copy of the prophage. A prophage can be replicated repeatedly in this way without killing the host cell or producing new phage particles. As mentioned earlier, this is called the lysogenic cycle.

Many bacteriophages can alternate between lysogenic and lytic cycles (Figure 10.3). A bacteriophage that may spend some of its time in the lysogenic cycle is called a **temperate phage.** Phage λ is an example of a temperate phage. Upon infection, it can either enter the lysogenic cycle or proceed directly to the lytic cycle. Other phages, called **virulent phages,** have only lytic cycles. The genome of a virulent phage is not capable of integration into a host chromosome. Phage T4, which we will examine later in this chapter, is a virulent phage that infects *E. coli*. Unlike phage λ, which may coexist harmlessly with *E. coli*, T4 always lyses the infected cell.

For phages such as λ that can follow either cycle, environmental conditions influence whether or not viral DNA is integrated into a host chromosome and how long the virus remains in the lysogenic cycle. If nutrients are readily available, phage λ usually proceeds directly to the lytic cycle after its DNA enters the cell. Alternatively, if nutrients are in short supply, the lysogenic cycle is often favored because sufficient material may not be available to make new viruses. If more nutrients become available later, this may cause the prophage to become activated. At this point, the viral reproductive cycle switches to the lytic cycle, and new viruses are made and released.

Latency in Human Viruses Latency among human viruses can occur in two different ways. For HIV, latency occurs because the virus has integrated into the host genome and may remain dormant for long periods of time. Alternatively, the genomes of other viruses can exist as an **episome**—a genetic element that can replicate independently of the chromosomal DNA but also can occasionally integrate into chromosomal DNA. Examples of

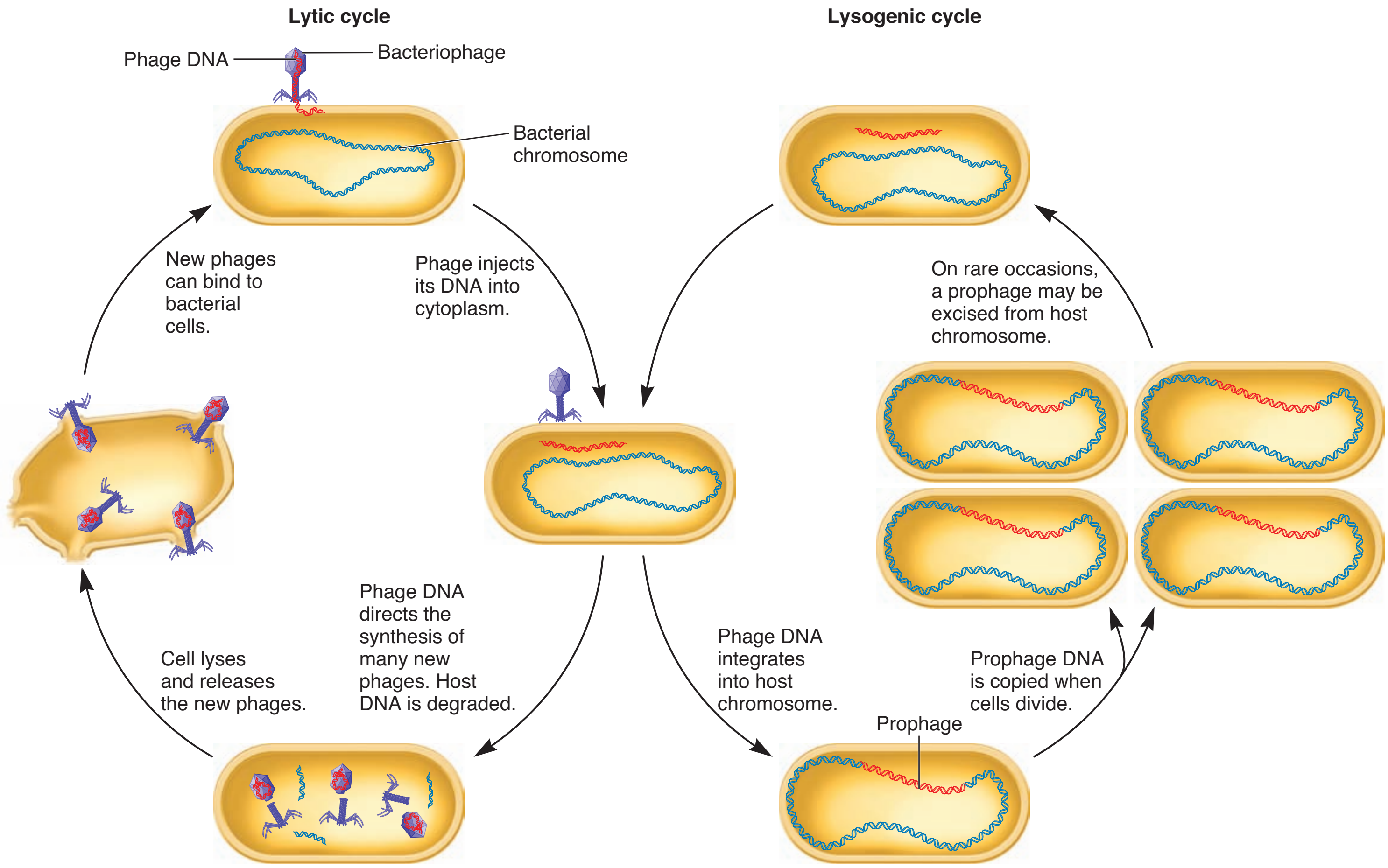

Lytic cycle
Phage DNA
Bacteriophage
Bacterial chromosome
Phage injects its DNA into cytoplasm.
Phage DNA directs the synthesis of many new phages. Host DNA is degraded.
Cell lyses and releases the new phages.
New phages can bind to bacterial cells.
Lysogenic cycle
Phage DNA integrates into host chromosome.
Prophage
Prophage DNA is copied when cells divide.
On rare occasions, a prophage may be excised from host chromosome.

FIGURE 10.3 **The lytic and lysogenic reproductive cycles of certain bacteriophages.** Some bacteriophages, such as temperate phages, can follow both cycles. Other phages, known as virulent phages, can follow only a lytic cycle.

Concept check: Which cycle produces new phage particles?

viral genomes that can exist as episomes include different types of herpesviruses that cause cold sores (herpes simplex type I), genital herpes (herpes simplex type II), and chickenpox (varicella-zoster virus). A person infected with a given type of herpesvirus may have periodic outbreaks of disease symptoms when the virus switches from the latent, episomal form to the active form that produces new virus particles.

As an example, let's consider the herpesvirus called varicella-zoster virus. The initial infection by this virus causes chickenpox, after which the virus may remain latent for many years as an episome. The disease called shingles occurs when varicella-zoster virus switches from the latent state and starts making new virus particles. It usually occurs in adults over the age of 60 or in people with weakened immune systems. Shingles begins as a painful rash that eventually erupts into blisters. The blisters follow the path of the nerve cells that carry the latent varicella-zoster virus. The blisters often form a ring around the back of the patient's body, which is why the disease is called shingles—the pattern of blisters line up like the shingles on a roof.

Emerging Viruses, Such as HIV, Have Arisen Recently and May Rapidly Spread Through a Population

A primary reason researchers have been interested in viral reproductive cycles is the ability of many viruses to cause diseases in humans and other hosts. Some examples of human disease-causing viruses were presented earlier in Table 10.1. **Emerging viruses** are viruses that have arisen recently and are more likely to cause infection compared with previous strains. Such viruses may lead to a significant loss of human life and often cause public alarm. New strains of influenza virus arise fairly regularly due to new mutations. An example is the strain H1N1, also called swine flu (Figure 10.4). It was originally called swine flu because laboratory testing revealed that it carries two genes that are normally found in flu viruses that infect pigs in Europe and Asia. The Centers for Disease Control and Prevention (CDC) recommends vaccination as the first and most important step in preventing infection by H1N1 and other strains of influenza.

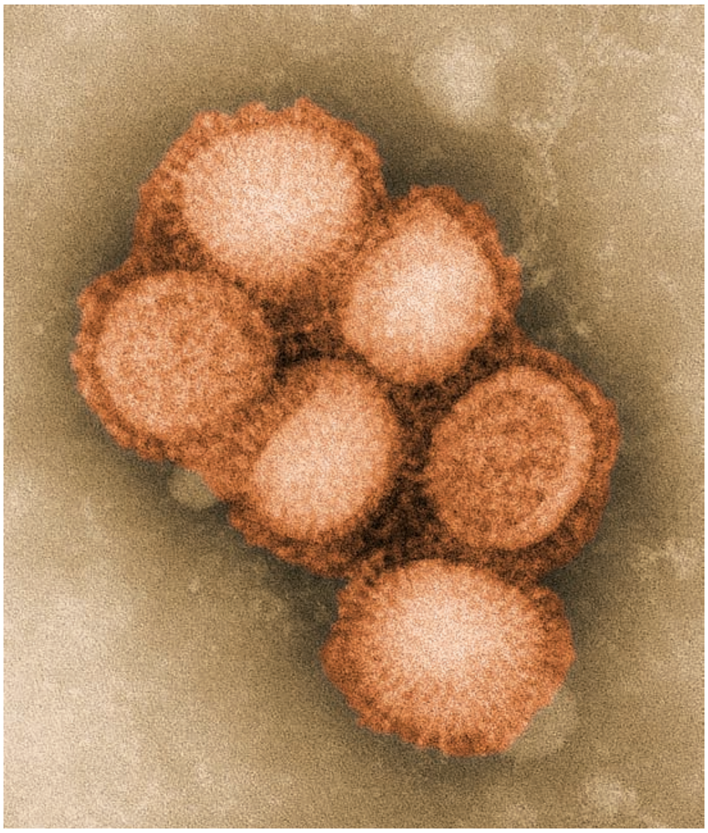


FIGURE 10.4 **A micrograph of influenza virus, strain H1N1.** This virus, which infects people, is a relatively new strain of influenza virus that causes classic flu symptoms.

However, in the United States, despite vaccination campaigns that aim to minimize deaths from seasonal influenza, over 30,000 people die annually from this disease.

Another emerging virus caused a potentially life-threatening illness called severe acute respiratory syndrome (SARS). This RNA virus was first identified by researchers in Hong Kong, the United States, and Germany in 2003. The type of virus that caused SARS is called a coronavirus. The SARS coronavirus is believed to have originated in bats and acquired the ability to infect humans and livestock. Symptoms in people included a high fever, headache, a dry cough, and pneumonia. The SARS outbreak ended in 2004.

During the past few decades, the most devastating example of an emerging virus has been human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS). HIV is primarily spread by sexual contact between infected and uninfected individuals, but it can also be spread by the transfusion of HIV-infected blood, by the sharing of needles among drug users, and from infected mother to unborn child. The total number of AIDS deaths between 1981 and the end of 2006 was over 25 million; more than 0.5 million of these deaths occurred in the United States. During 2008, around 3 million adults and children became infected with HIV worldwide; nearly 1 in every 100 adults between ages 15 and 49 is infected. In the United States, about 55,000 new HIV infections occur each year, 70% of which are in men and 30% in women.

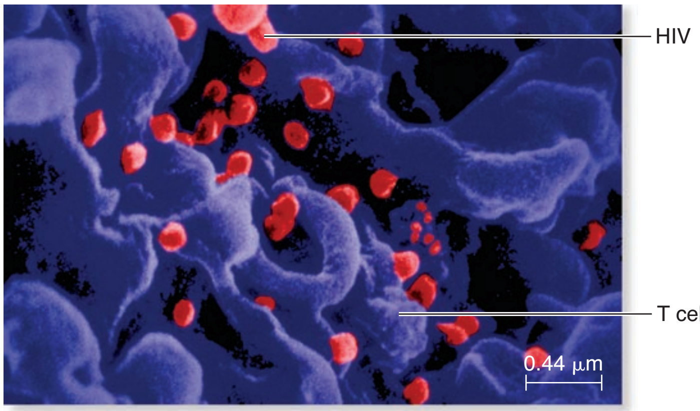


FIGURE 10.5 **Micrograph of HIV invading a human helper T cell.** This is a colorized scanning electron micrograph. The surface of the T cell is purple, and HIV particles are red.

The devastating effects of AIDS result from viral destruction of helper T cells, a type of white blood cell that plays an essential role in the immune system of mammals. **Figure 10.5** shows HIV virus particles invading a helper T cell, which interact with other cells of the immune system to facilitate the production of antibodies and other molecules that target and kill foreign invaders of the body. When large numbers of helper T cells are destroyed by HIV, the function of the immune system is seriously compromised, and the individual becomes susceptible to opportunistic infections, infectious diseases that would not normally occur in a healthy person. For example, *Pneumocystis jiroveci*, a fungus that causes pneumonia, is easily destroyed by a normal immune system. However, in people with AIDS, infection by this fungus can be fatal. An insidious feature of HIV replication, described in Figure 10.2b, is that reverse transcriptase, the enzyme that copies the RNA genome into DNA, lacks a proofreading function. In Chapter 13, we will discuss how DNA polymerase can identify and remove mismatched nucleotides in newly synthesized DNA. Because reverse transcriptase lacks this function, it makes more errors and thereby tends to create mutant strains of HIV. This undermines the ability of the body to combat HIV because mutant strains may not be destroyed by the body's defenses. In addition, mutant strains of HIV may be resistant to antiviral drugs.

10.2 REVIEWING THE KEY CONCEPTS

- The viral reproductive cycle consists of five or six basic steps, including attachment, entry, integration, synthesis, assembly, and release (see Figure 10.2).
- Some bacteriophages can alternate between two reproductive cycles: the lytic cycle and the lysogenic cycle (see Figure 10.3).
- Emerging viruses have arisen recently and are more likely to cause infection (see Figure 10.4).
- The disease AIDS is caused by a virus called human immunodeficiency virus (HIV). The virus is a retrovirus whose reproductive cycle involves the integration of the viral genome into a chromosome in the host cell (see Figure 10.5).

10.2 COMPREHENSION QUESTIONS

1. Which of the following is a common order of steps for a viral reproductive cycle? Note: integration is an optional step.
   a. Entry, integration, attachment, synthesis of viral components, viral assembly, viral release
   b. Entry, integration, synthesis of viral components, viral assembly, attachment, viral release
   c. Attachment, entry, integration, synthesis of viral components, viral assembly, viral release
   d. Attachment, entry, integration, viral assembly, synthesis of viral components, viral release
2. An example of an emerging virus is
   a. a coronavirus that causes SARS.
   b. HIV that causes AIDS.
   c. a strain of influenza virus called H1N1 that causes the flu.
   d. all of the above.

10.3 PLAQUE FORMATION AND INTERGENIC COMPLEMENTATION IN BACTERIOPHAGES

In this section and the one following, we will focus our attention on a bacteriophage called T4. Its genetic material contains several dozen genes encoding different proteins that carry out a variety of functions. For example, some of the genes encode proteins needed for the synthesis of new viruses and the lysis of the host cell. Other genes encode the viral coat proteins that are found in the head, shaft, base plate, and tail fibers. **Figure 10.6** illustrates the structure of T4. As seen here, five different proteins bind to each other to form a tail fiber. The expression of T4 genes to make these proteins provides the bacteriophage with the trait of having tail fibers that enable it to attach to the surface of a bacterium.

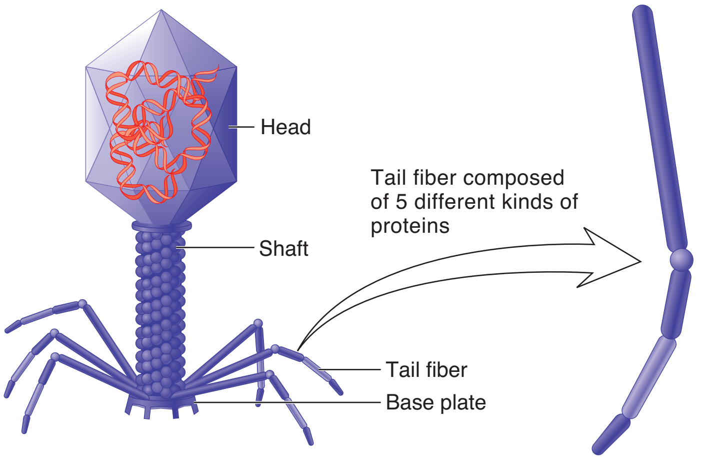


FIGURE 10.6 **Structure of the T4 virus.**

Genes → Traits The inset to this figure shows the five different proteins, encoded by five different genes, that make up a tail fiber. The expression of these genes provides the bacteriophage with the trait of having tail fibers, enabling it to attach itself to the surface of a bacterium.

Concept check: Where is the genetic material located that determines the structure of the virus?

The study of viral genes has been instrumental in our basic understanding of how the genetic material works. During the 1950s, Seymour Benzer embarked on a 10-year study that focused on the function of viral genes in the T4 bacteriophage. In this section, we will examine the growth properties of bacteriophages in the laboratory and explore how mutations in bacteriophage genes are analyzed.

Lytic Bacteriophages Form Plaques within a Lawn of Bacterial Cells

As they progress through the lytic cycle, bacteriophages ultimately produce new phages, which are released when the bacterial cell lyses (refer back to Figure 10.3, left side). In the laboratory, researchers can visually observe the consequences of bacterial cell lysis in the following way (**Figure 10.7**). In this example, a sample of many bacterial cells and a few lytic bacteriophages were mixed together and then poured onto petri plates containing nutrient agar for bacterial cell growth. Most bacterial cells were not infected by a bacteriophage and rapidly grew and divided to produce a continuous layer or lawn of bacteria. The lawn is opaque—you cannot see through it to the underlying agar. Each bacteriophage infected a bacterial cell found at random locations in the lawn of uninfected bacteria. The infected cells lysed and released newly made bacteriophages. These newly made bacteriophages then infected the nearby bacteria within the lawn. These cells eventually lysed and also released newly made phages. As shown in the third plate from the top, these repeated cycles of infection and lysis produced an observable clear area, or **plaque,** where the bacteria had been lysed around the original site where a phage infected a bacterial cell. In this case, 11 bacteriophages initially infected *E. coli* cells and produced 11 distinct plaques within a lawn of uninfected cells.

Mutations in Viral Genes Can Alter Plaque Morphology

Now that we have an appreciation for the formation of a viral plaque, let's consider how the characteristics of a plaque can be viewed as a trait of a bacteriophage. As we have seen, the genetic analysis of any organism requires strains with allelic differences. Because bacteriophages can be visualized only with an electron microscope, it is rather difficult for geneticists to analyze mutations that affect phage morphology. However, some mutations in the bacteriophage's genetic material alter the ability of the phage to cause plaque formation. Therefore, we can view the morphology of plaques as a trait of the bacteriophage. Because plaques are visible with the unaided eye, mutations affecting this trait lend themselves to a much easier genetic analysis. An example is a rapid-lysis mutant of bacteriophage T4, which tends to form unusually large plaques (**Figure 10.8**). The plaques are large because the mutant phages lyse the bacterial cells more quickly than do the wild-type phages. Rapid-lysis mutants form large, clearly defined plaques, as opposed to wild-type bacteriophages, which produce smaller,

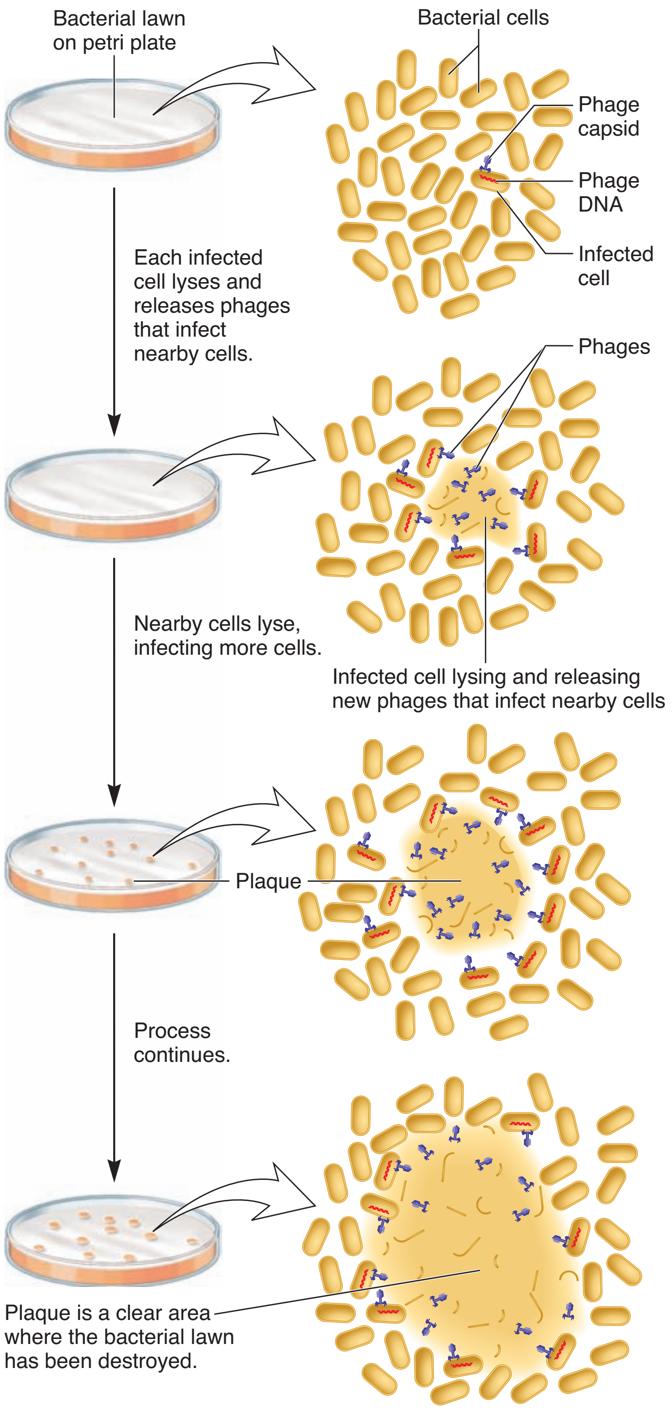
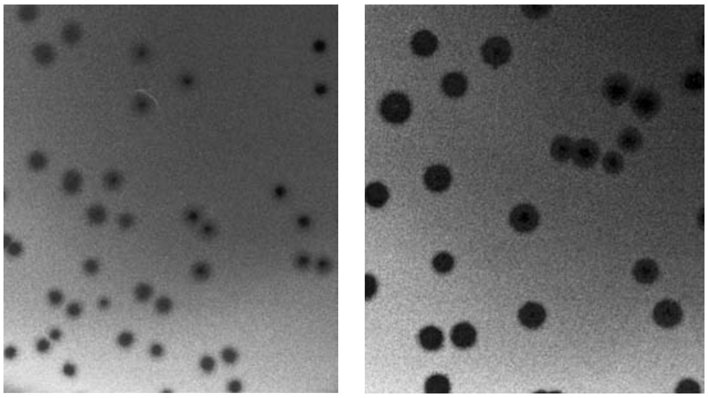

Bacterial lawn on petri plate
Bacterial cells
Phage capsid
Phage DNA
Infected cell
Each infected cell lyses and releases phages that infect nearby cells.
Phages
Nearby cells lyse, infecting more cells.
Infected cell lysing and releasing new phages that infect nearby cells
Plaque
Process continues.
Plaque is a clear area where the bacterial lawn has been destroyed.

FIGURE 10.7 **The formation of phage plaques on a lawn of bacteria in the laboratory.** In this experiment, bacterial cells were mixed with a small number of lytic bacteriophages. In this particular example, 11 bacterial cells were initially infected by phages. The cells were poured onto petri plates containing nutrient agar for bacterial cell growth. Bacterial cells rapidly grow and divide to produce an opaque lawn of densely packed bacteria. The 11 infected cells lyse and release newly made bacteriophages. These bacteriophages then infect the nearby bacteria within the lawn. Likewise, these newly infected cells lyse and release new phages. By this repeated process, the area of cell lysis creates a clear zone known as a viral plaque.

Concept check: Over time, the plaques become larger in diameter. Explain why.



(a) Plaques caused by wild-type bacteriophages **(b) Plaques caused by rapid-lysis bacteriophage strains**

FIGURE 10.8 **A comparison of plaques produced by the wild-type T4 bacteriophage and a rapid-lysis mutant.**

Genes → Traits **(a)** These plaques were caused by the infection and lysis of *E. coli* cells by the wild-type T4 phage. **(b)** A mutation in a phage gene, called a rapid-lysis mutation, caused the phage to lyse *E. coli* cells more quickly. A phage carrying a rapid-lysis mutant allele yields much larger plaques with clearly defined edges.

fuzzy-edged plaques. Mutations in different bacteriophage genes can produce a rapid-lysis phenotype.

Benzer studied one category of T4 phage mutants, designated *rII* (*r* stands for rapid lysis). In the bacterial strain called *E. coli* B, *rII* phages produce abnormally large plaques. Nevertheless, *E. coli* B strains produce low yields of *rII* phages because the *rII* phages lyse the bacterial cells so quickly they do not have sufficient time to produce many new phages. To help study this phage, Benzer wanted to obtain large quantities of it. Therefore, to improve his yield of *rII* phage, he decided to test its yield in other bacterial strains.

On the day Benzer decided to do this, he happened to be teaching a phage genetics class. For that class, he was growing two *E. coli* strains designated *E. coli* K12S and *E. coli* K12(λ). He was growing these two strains to teach his class about the lysogenic cycle. *E. coli* K12(λ) has DNA from another phage, called lambda, integrated into its chromosome, whereas *E. coli* K12S does not. To see if the use of these strains might improve phage yield, *E. coli* B, *E. coli* K12S, and *E. coli* K12(λ) were infected with the *rII* and wild-type T4 phage strains. As expected, the wild-type phage could infect all three bacterial strains. However, the *rII* mutant strains behaved quite differently. In *E. coli* B, the *rII* strains produced large plaques that had poor yields of bacteriophage. In *E. coli* K12S, the *rII* mutants produced normal plaques that gave good yields of phage. Surprisingly, in *E. coli* K12(λ), the *rII* mutants were unable to produce plaques at all, for reasons that were not understood at the time. Nevertheless, as we will see later, this fortuitous observation was a critical feature that allowed intragenic mapping in this bacteriophage.

A Complementation Test Can Reveal If Mutations Are in the Same Gene or in Different Genes

In his experiments, Benzer was interested in a single trait, namely, the ability to form plaques. He had isolated many *rII* mutant strains that could form large plaques in *E. coli* B but could not produce plaques in *E. coli* K12(λ). To attempt gene mapping, he needed to know if the various *rII* mutations were in the same gene or if they involved mutations in different phage genes. To accomplish this, he conducted a **complementation test.** The goal of this type of approach is to determine if two different mutations that affect the same trait are in the same gene or in two different genes.

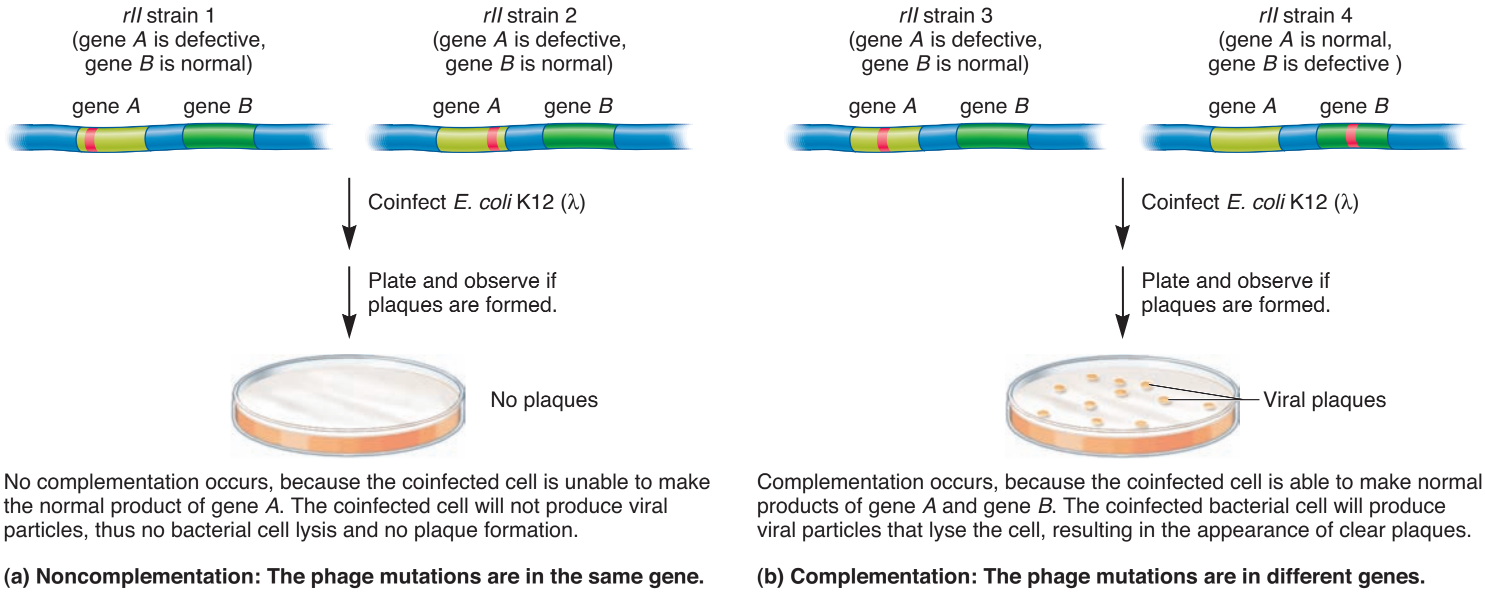
The possible outcomes of complementation tests involving mutations that affect plaque formation are shown in **Figure 10.9**. This example involves four different *rII* mutations in T4 bacteriophage, designated strains 1 through 4, that prevent plaque formation in *E. coli* K12(λ). To conduct this complementation experiment, bacterial cells were coinfected with an excess of two different strains of T4 phage. Two distinct outcomes are possible. In Figure 10.9a, the two *rII* phage strains have mutations in the same gene (gene *A*). Because they cannot make a wild-type gene *A* product when coinfected into an *E. coli* K12(λ) cell, plaques do not form. This phenomenon is called **noncomplementation.**

Alternatively, if each *rII* mutation is in a different phage gene (e.g., gene *A* and gene *B*), a bacterial cell that is coinfected by both types of phages will have two mutant genes as well as two wild-type genes (Figure 10.9b). If the mutant phage genes behave in a recessive fashion, the doubly infected cell will have a wild-type phenotype. Why does this phenotype occur? The coinfected cells produce normal proteins that are encoded by the wild-type versions of both genes *A* and *B*. For this reason, coinfected cells are lysed in the same manner as if they were infected by the wild-type strain. Therefore, this coinfection is able to produce plaques in *E. coli* K12(λ). This result is called **complementation** because the defective genes in each *rII* strain are complemented by the corresponding wild-type genes. It should be noted that, for a variety of reasons, intergenic complementation may not always work. One possibility is that a mutation may behave in a dominant fashion. In addition, mutations that affect regulatory genetic regions rather than the protein-coding region may not show complementation.

By carefully considering the pattern of complementation and noncomplementation, Benzer found that the *rII* mutations occurred in two different genes, which were termed *rIIA* and *rIIB*. The identification of two distinct genes affecting plaque formation was a necessary step that preceded his intragenic mapping analysis, which is described next. Benzer coined the term **cistron** to refer to the smallest genetic unit that gives a negative complementation test. In other words, if two mutations occur within the same cistron, they cannot complement each other. Since these studies, researchers learned that a cistron is equivalent to a gene. In recent decades, the term *gene* has gained wide popularity, and the term *cistron* is not commonly used. However, the term *polycistronic* is still used to describe bacterial RNAs that carry two or more gene sequences, as described in Chapter 16.

10.3 REVIEWING THE KEY CONCEPTS

- Bacteriophages are viruses that infect bacteria (see Figure 10.6).
- Some bacteriophages cause bacteria to lyse. This event can lead to plaque formation when bacteria are plated as a lawn on bacterial media (see Figure 10.7).
- Studying a type of bacteriophage called T4, Benzer identified mutant strains that caused rapid lysis, thereby leading to larger plaques (see Figure 10.8).



No complementation occurs, because the coinfected cell is unable to make the normal product of gene *A*. The coinfected cell will not produce viral particles, thus no bacterial cell lysis and no plaque formation.

(a) Noncomplementation: The phage mutations are in the same gene.

Complementation occurs, because the coinfected cell is able to make normal products of gene *A* and gene *B*. The coinfected bacterial cell will produce viral particles that lyse the cell, resulting in the appearance of clear plaques.

(b) Complementation: The phage mutations are in different genes.

FIGURE 10.9 **A comparison of noncomplementation and complementation.** Four different T4 phage strains (designated 1 through 4) that carry *rII* mutations were coinfected into *E. coli* K12(λ). **(a)** If two *rII* phage strains possess mutations in the same gene, noncomplementation will occur. **(b)** If the *rII* mutations are in different genes (such as gene *A* and gene *B*), a coinfected cell will have two mutant genes but also two wild-type genes. Doubly infected cells with a wild-type copy of each gene can produce new phages and form plaques. This result is called complementation because the defective genes in each *rII* strain are complemented by the corresponding wild-type genes.

Concept check: *With regard to the locations of mutations, explain the difference between noncomplementation and complementation.*

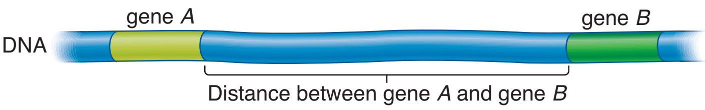
- Benzer identified strains of T4 that could not form plaques in *E. coli* K12(λ). He conducted complementation tests to determine if the mutations in these strains were in the same phage gene or in different genes (see Figure 10.9).

10.3 COMPREHENSION QUESTIONS

1. A bacterial plaque is
   a. a clear region in a bacterial lawn where the bacteria have been lysed.
   b. a bacterial cell that has a phage attached to its surface.
   c. a dense region of bacterial growth.
   d. none of the above.
2. When two different alleles that affect the same trait exhibit complementation, an interpretation is that
   a. the alleles occur in the same gene.
   b. the alleles occur in two different genes.
   c. both alleles are dominant.
   d. one allele is dominant and the other is recessive.

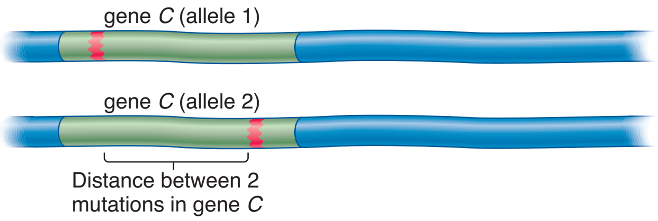
10.4 INTRAGENIC MAPPING IN BACTERIOPHAGES

In the previous section, we learned how Benzer was able to identify T4 bacteriophage mutations that altered plaque morphology and to determine if such mutations were in the same or different genes via complementation. In this section, we will examine how he conducted a detailed type of genetic mapping known as **intragenic mapping,** or **fine-structure mapping.** How is this mapping different from mapping studies in eukaryotes and bacteria, which we considered in Chapters 7 and 9, respectively? The mapping studies in eukaryotes and bacteria are examples of **intergenic mapping.** The goal of intergenic mapping is to determine the distance between two different genes. For example, the determination of the distance between genes *A* and *B* is an example of intergenic mapping:



Intergenic mapping

In comparison, intragenic mapping seeks to establish distances between two or more mutations within the same gene. For example, in a population of viruses, gene *C* may exist as two different mutant alleles. One allele may be due to a mutation near the beginning of the gene; the second, to a mutation near the end:



Intragenic mapping (also known as fine-structure mapping)

In this section, we will explore the pioneering studies that led to the development of intragenic mapping and advanced our knowledge of gene function. Benzer's results showed that, rather than being an indivisible particle, a gene must be composed of a large structure that can be subdivided during intragenic crossing over.

Intragenic Maps Were Constructed Using Data from a Recombinational Analysis of Mutants Within the *rII* Region

As we saw in Figure 10.9, the ability of strains with mutations in two different genes to produce viral plaques after coinfection is due to complementation. Noncomplementation occurs when two different strains have mutations in the same gene. However, at an extremely low rate, two noncomplementing strains of viruses can produce an occasional viral plaque if intragenic recombination has taken place. For example, **Figure 10.10** shows a coinfection experiment between two phage strains that both contain *rII* mutations in gene *A*. These mutations are located at different places within the same gene. On very rare occasions, a crossover may occur in the short region between each mutation. This crossover produces a double-mutant gene *A* and a wild-type gene *A*. Because this event has produced a wild-type gene *A*, the function of the protein encoded by gene *A* is restored. Therefore, new phages can be made in *E. coli* K12(λ), resulting in the formation of a viral plaque.

Figure 10.11 describes the general strategy for intragenic mapping of *rII* phage mutations. Bacteriophages from two different noncomplementing *rII* phage mutants (here, *r103* and *r104*) were mixed together in equal numbers and then infected into

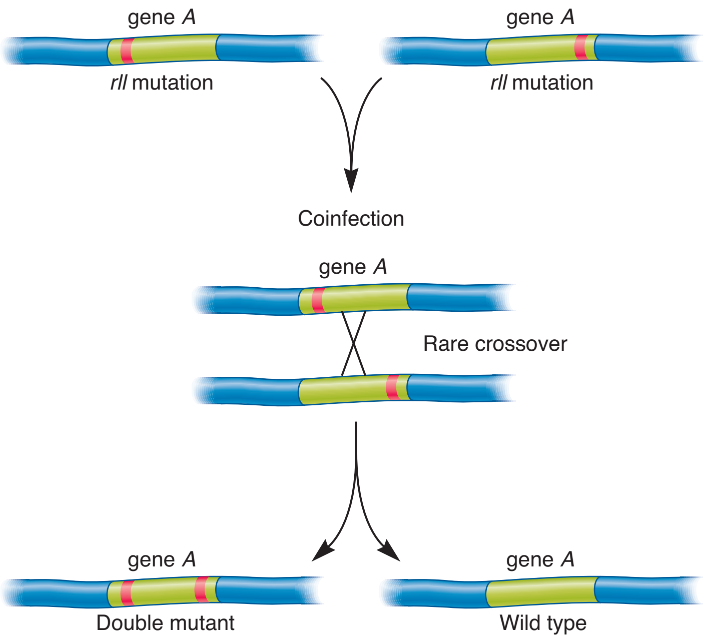
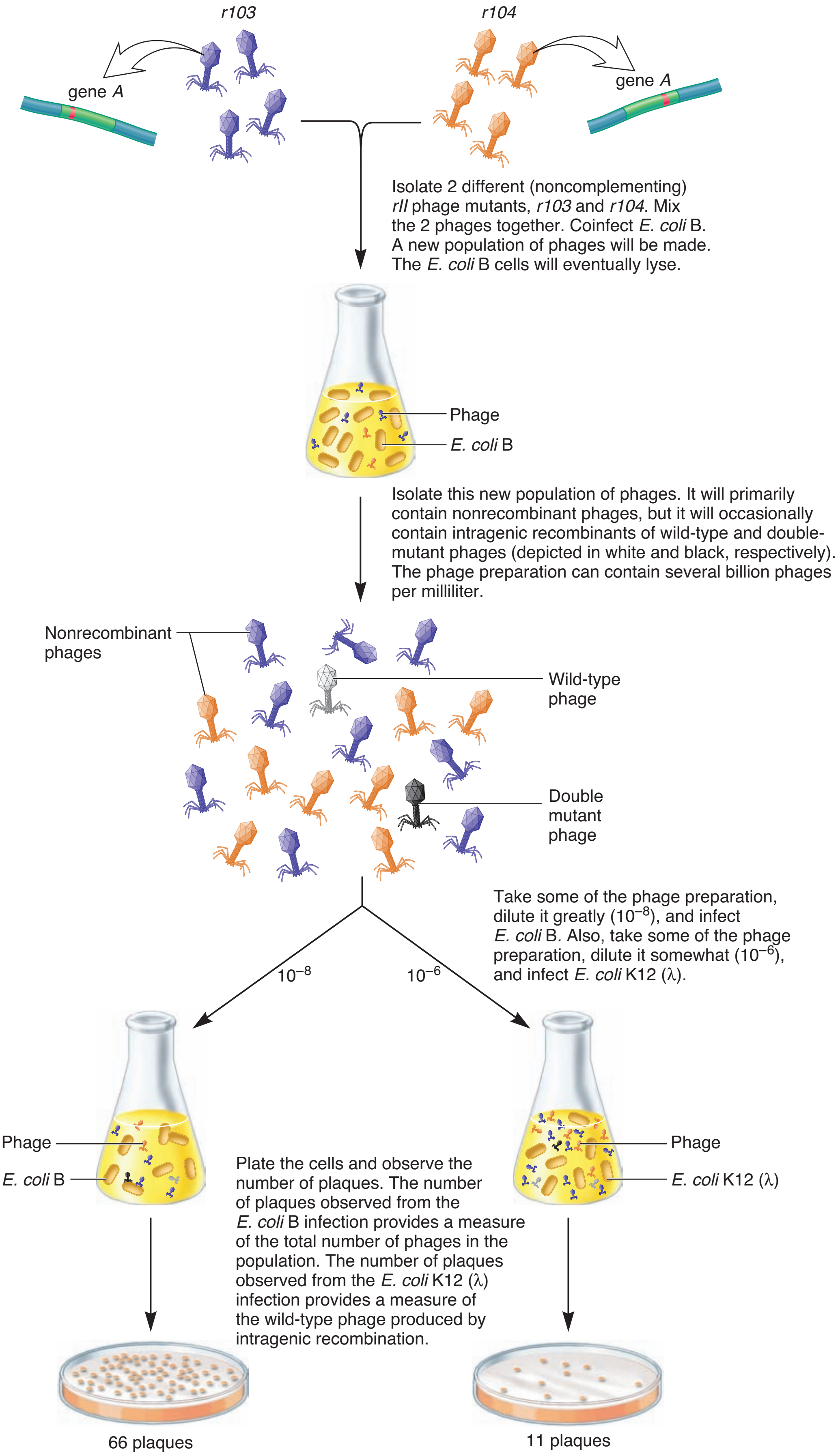


FIGURE 10.10 **Intragenic recombination.** Following coinfection, a rare crossover has occurred between the sites of the two mutations. This produces a wild-type phage with no mutations and a double-mutant phage with both mutations.


r103
r104
gene *A*
gene *A*
Phage
E. coli B
Nonrecombinant phages
Wild-type phage
Double mutant phage
10^{-8}
10^{-6}
Phage
E. coli B
Phage
E. coli K12 (λ)
66 plaques
11 plaques

Isolate 2 different (noncomplementing) *rII* phage mutants, *r103* and *r104*. Mix the 2 phages together. Coinfect *E. coli* B. A new population of phages will be made. The *E. coli* B cells will eventually lyse.

Isolate this new population of phages. It will primarily contain nonrecombinant phages, but it will occasionally contain intragenic recombinants of wild-type and double-mutant phages (depicted in white and black, respectively). The phage preparation can contain several billion phages per milliliter.

Take some of the phage preparation, dilute it greatly (10^{-8}), and infect *E. coli* B. Also, take some of the phage preparation, dilute it somewhat (10^{-6}), and infect *E. coli* K12 (λ).

Plate the cells and observe the number of plaques. The number of plaques observed from the *E. coli* B infection provides a measure of the total number of phages in the population. The number of plaques observed from the *E. coli* K12 (λ) infection provides a measure of the wild-type phage produced by intragenic recombination.

FIGURE 10.11 **Benzer's method of intragenic mapping in the *rII* region.**

Concept check: Why was it necessary to infect E. coli K12(λ) with 100 times more phage compared to E. coli B?

E. coli B, and a new population of phages was isolated. This new population was expected to contain predominantly nonrecombinant phages. However, due to intragenic recombination, it was also expected to contain a very low percentage of wild-type phages and double-mutant phages (refer back to Figure 10.10).

How could Benzer determine the number of rare phages that were produced by intragenic recombination? The key approach is that *rII* mutant phages cannot grow in *E. coli* K12(λ). Following coinfection, he took this new population of phages and used some of them to infect *E. coli* B and some to infect *E. coli* K12(λ). After plating, the *E. coli* B infection was used to determine the total number of phages, because *rII* mutants as well as wild-type phages can produce plaques in this strain. The overwhelming majority of these phages were expected to be nonrecombinant phages. The *E. coli* K12(λ) infection was used to determine the number of rare intragenic recombinants that produce wild-type phages.

Figure 10.11 illustrates the great advantage of this experimental system in detecting a low percentage of recombinants. In the laboratory, phage preparations containing several billion phages per milliliter are readily made. Among billions of phages, a low percentage (e.g., 1 in every 1000) may be wild-type phages arising from intragenic recombination. The wild-type recombinants can produce plaques in *E. coli* K12(λ), whereas the *rII* mutant strains cannot. In other words, only the tiny fraction of wild-type recombinants produce plaques in *E. coli* K12(λ).

The frequency of recombinant phages can be determined by comparing the number of wild-type phages, produced by intragenic recombination, and the total number of phages. As shown in Figure 10.11, the total number of phages can be deduced from the number of plaques obtained from the infection of *E. coli* B. In this experiment, the phage preparation was diluted by 10^8 (1:100,000,000), and 1 mL was used to infect *E. coli* B. Because this plate produced 66 plaques, the total number of phages in the original preparation was $66 \times 10^8 = 6.6 \times 10^9$, or 6.6 billion phages per milliliter. By comparison, the phage preparation used to infect *E. coli* K12(λ) was diluted by only 10^6 (1 : 1,000,000). This plate produced 11 plaques. Therefore, the number of wild-type phages was 11×10^6, which equals 11 million wild-type phages per milliliter.

As we have already seen in Chapter 7, genetic mapping distance is computed by dividing the number of recombinants by the total population (nonrecombinants and recombinants) times 100. In this experiment, intragenic recombination produced an equal number of two types of recombinants: wild-type phages and double-mutant phages. Only the wild-type phages are detected in the infection of *E. coli* K12(λ). Therefore, to obtain the total number of recombinants, the number of wild-type phages must be multiplied by 2. With all of this information, we can use the following equation to compute the frequency of recombinants using the experimental approach described in Figure 10.11.

$$\text{Frequency of recombinants} = \frac{2\ [\text{Wild-type plaques obtained in } \textit{E. coli}\ \text{K12}(\lambda)]}{\text{Total number of plaques obtained in } \textit{E. coli}\ \text{B}}$$

$$\text{Frequency of recombinants} = \frac{2\ (11 \times 10^6)}{6.6 \times 10^9}$$

$$= 3.3 \times 10^{-3} = 0.0033$$

In this example, approximately 3.3 recombinants were produced per 1000 phages.

The frequency of recombinants provides a measure of map distance. In eukaryotic mapping studies, we compute the map distance by multiplying the frequency of recombinants by 100 to give a value in map units (also known as centiMorgans). Similarly, in these experiments, the frequency of recombinants can provide a measure of map distance along the bacteriophage DNA. In this case, the map distance is between two mutations within the same gene. Like intergenic mapping, the frequency of intragenic recombinants is correlated with the distance between the two mutations; the farther apart they are, the higher the frequency of recombinants. If two mutations happen to be located at exactly the same site within a gene, coinfection would not be able to produce any wild-type recombinants, and so the map distance would be zero. These are known as **homoallelic** mutations.

Deletion Mapping Can Be Used to Localize Many *rII* Mutations to Specific Regions in the *rIIA* or *rIIB* Genes

Now that we have seen the general approach to intragenic mapping, let's consider a method to efficiently map hundreds of *rII* mutations within the two genes designated *rIIA* and *rIIB*. As you may have realized, the coinfection experiments described in Figure 10.11 are quite similar to Sturtevant's strategy of making dihybrid crosses to map genes along the X chromosome of *Drosophila* (refer back to Figure 7.9). Similarly, Benzer wanted to coinfect different *rII* mutants in order to map the sites of the mutations within the *rIIA* and *rIIB* genes. During the course of his work, he obtained hundreds of different *rII* mutant strains that he wanted to map. However, making all of the pairwise combinations would have been an overwhelming task. Instead, Benzer used an approach known as **deletion mapping** as a first step in localizing his *rII* mutations to a fairly short region within gene *A* or gene *B*.

Figure 10.12 describes the general strategy used in deletion mapping. This approach is easier to understand if we use an example. Let's suppose that the goal is to know the approximate location of an *rII* mutation, such as *r103*. To do so, *E. coli* K12(λ) is coinfected with *r103* and a deletion strain. Each deletion strain is a T4 bacteriophage that is missing a known segment of the *rIIA* and/or *rIIB* gene. If the deleted region includes the same region that contains the *r103* mutation, a coinfection cannot produce intragenic wild-type recombinants. Therefore, plaques cannot form. However, if a deletion strain recombines with *r103* to produce a wild-type phage, the deleted region does not overlap with the *r103* mutation.

In the example shown in Figure 10.12, the *r103* strain produced wild-type recombinants when coinfected with deletion strains *PB242*, *A105*, and *638*. However, coinfection of *r103* with *PT1*, *J3*, *1241*, and *1272* did not produce intragenic wild-type recombinants. Because coinfection with *PB242* produced recombinants and *PT1* did not, the *r103* mutation must be located in the region that is missing in *PT1* but not missing in *PB242*. As shown at the bottom of Figure 10.12, this region is called A4 (the A refers to the *rIIA* gene). In other words, the *r103* mutation is

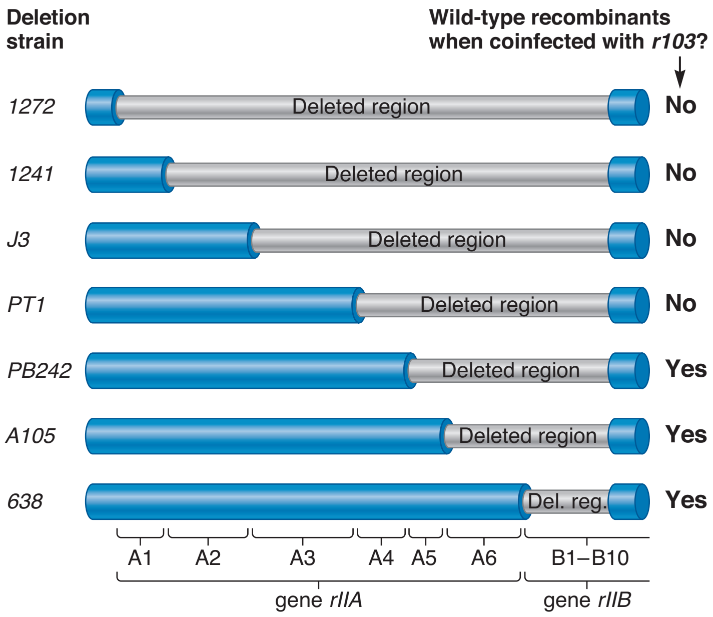

Deletion strain
Wild-type recombinants when coinfected with *r103*?
1272
Deleted region
No
1241
Deleted region
No
J3
Deleted region
No
PT1
Deleted region
No
PB242
Deleted region
Yes
A105
Deleted region
Yes
638
Del. reg.
Yes
A1
A2
A3
A4
A5
A6
B1–B10
gene *rIIA*
gene *rIIB*

FIGURE 10.12 The use of deletion strains to localize *rII* mutants to short regions within the *rIIA* or *rIIB* gene. The deleted regions are shown in gray.

Concept check: What is the practical benefit of deletion mapping?

located somewhere within the A4 region, but not in the other six regions (A1, A2, A3, A5, A6, and B).

As described in Figure 10.12, this first step in the deletion mapping strategy localized an *rII* mutation to one of seven regions; six of these were in *rIIA* and one was in *rIIB*. Other deletion strains were used to eventually localize each *rII* mutation to one of 47 short regions; 36 were in *rIIA*, 11 in *rIIB*. At this point, pairwise coinfections were made between mutant strains that had been localized to the same region by deletion mapping. For example, 24 mutations were deletion-mapped to a region called A5d. Pairwise coinfection experiments were conducted among this group of 24 mutants to precisely map their locations relative to each other in the A5d region. Similarly, all mutants in each of the 46 other groups were mapped by pairwise coinfections. In this way, a fine-structure map was constructed depicting the locations of hundreds of different *rII* mutations (**Figure 10.13**). As seen in this figure, certain locations contained a relatively high number of mutations compared with other sites. These were termed **hot spots** for mutation.

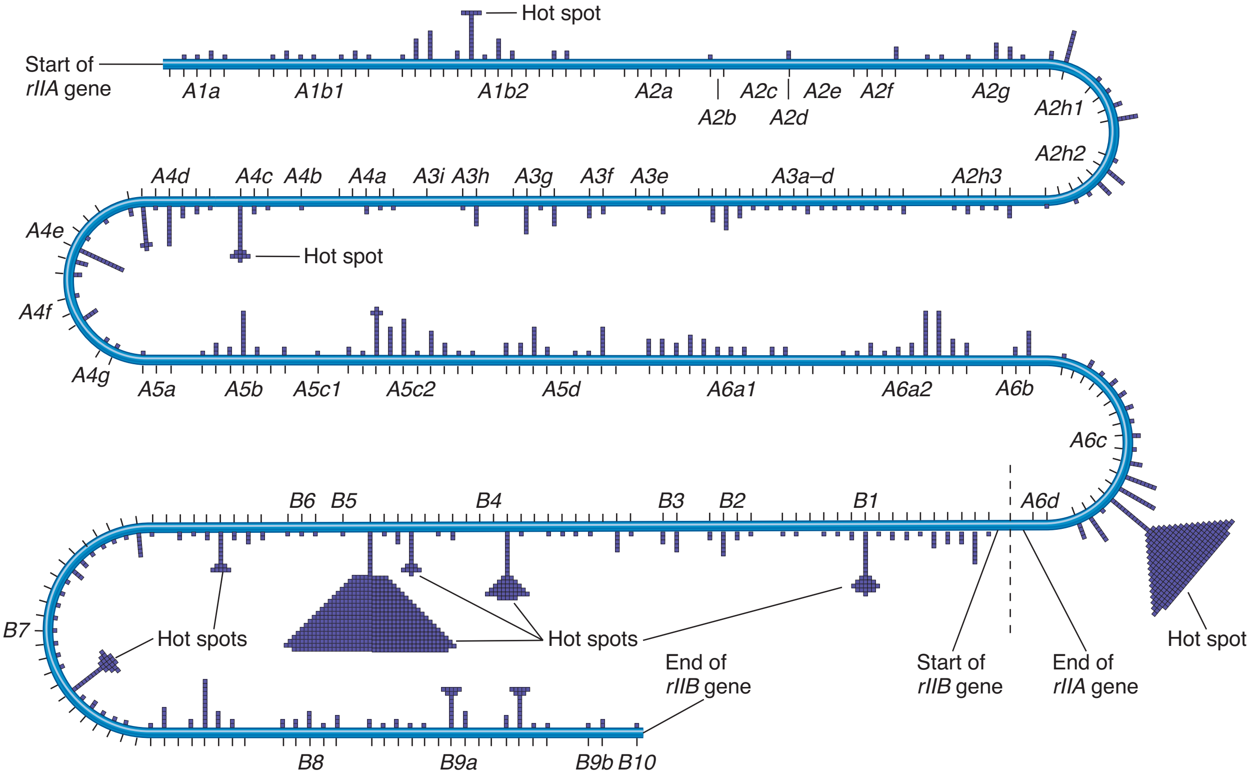

Hot spot
Start of *rIIA* gene
A1a
A1b1
A1b2
A2a
A2b
A2c
A2d
A2e
A2f
A2g
A2h1
A2h2
A2h3
A3a–d
A3e
A3f
A3g
A3h
A3i
A4a
A4b
A4c
A4d
Hot spot
A4e
A4f
A4g
A5a
A5b
A5c1
A5c2
A5d
A6a1
A6a2
A6b
A6c
A6d
B1
B2
B3
B4
B5
B6
B7
Hot spots
Hot spots
End of *rIIB* gene
Start of *rIIB* gene
End of *rIIA* gene
Hot spot
B8
B9a
B9b
B10

FIGURE 10.13 The outcome of intragenic mapping of many *rII* mutations. The blue line represents the linear sequence of the *rIIA* and *rIIB* genes, which are found within the T4 phage's genetic material. Each small purple box attached to the blue line symbolizes a mutation that was mapped by intragenic mapping. Among hundreds of independent mutant phages, several mutations sometimes mapped to the same site. In this figure, mutations at the same site form columns of boxes. Hot spots contain a large number of mutations and are represented as a group of boxes attached to a column of boxes. A hot spot contains many mutations at the same site within the *rIIA* or *rIIB* gene.

Concept check: In your own words, describe what this figure is illustrating.

Intragenic Mapping Experiments Provided Insight into the Relationship Between Traits and Molecular Genetics

Intragenic mapping studies were a pivotal achievement in our early understanding of gene structure. Since the time of Mendel, geneticists had considered a gene to be the smallest unit of heredity, which provided an organism with its inherited traits. In the late 1950s, however, the molecular nature of the gene was not understood. Because it is a unit of heredity, some scientists envisioned a gene as being a particle-like entity that could not be further subdivided into additional parts. However, intragenic mapping studies revealed, convincingly, that this is not the case. These studies showed that mutations can occur at many different sites within a single gene. Furthermore, intragenic crossing over could recombine these mutations, resulting in wild-type genes. Therefore, rather than being an indivisible particle, a gene must be composed of a large structure that can be subdivided during crossing over.

Benzer's results were published in the late 1950s and early 1960s, not long after the physical structure of DNA had been elucidated by Watson and Crick. We now know that a gene is a segment of DNA that is composed of smaller building blocks called nucleotides. A typical gene is a linear sequence of several hundred to many thousand base pairs. As the genetic map of Figure 10.13 indicates, mutations can occur at many sites along the linear structure of a gene; intragenic crossing over can recombine mutations that are located at different sites within the same gene.

10.4 REVIEWING THE KEY CONCEPTS

- Benzer devised a method to study intragenic recombination. He used deletion mapping to determine the locations of the mutations within two phage genes (see Figures 10.10–10.13).

10.4 COMPREHENSION QUESTIONS

1. The goal of intragenic mapping is to determine the
   a. order and distance between different genes.
   b. order and distance between mutations in a single gene.
   c. locations of genes within a viral genome.
   d. insertion site of a viral genome into the host cell's chromosome.
2. The purpose of deletion mapping is to
   a. determine if deletions affect the virulence of a virus.
   b. determine where a deletion is located within a particular gene.
   c. localize mutations to a defined region within a gene.
   d. determine if mutations are in the same gene or in different genes.

KEY TERMS

Page 207. virus, host cell, host range, capsid, capsomers, viral envelope

Page 209. bacteriophages, phages, viral genome, viral reproductive cycle, phage λ, human immunodeficiency virus (HIV), integrase, prophage, lysogenic cycle

Page 212. reverse transcriptase, provirus, lytic cycle, latent, lysogeny, temperate phage, virulent phages, episome

Page 213. emerging viruses

Page 215. plaque

Page 217. complementation test, noncomplementation, complementation, cistron

Page 218. intragenic mapping, fine-structure mapping, intergenic mapping

Page 220. homoallelic, deletion mapping

Page 221. hot spots

CHAPTER SUMMARY

10.1 Virus Structure and Genetic Composition

- Tobacco mosaic virus (TMV) was the first virus to be discovered. It infects many species of plants.
- Viruses vary with regard to their host range, structure, and genome composition (see Table 10.1, Figure 10.1).

10.2 Viral Reproductive Cycles

- The viral reproductive cycle consists of five or six basic steps, including attachment, entry, integration, synthesis, assembly, and release (see Figure 10.2).
- Some bacteriophages can alternate between two reproductive cycles: the lytic cycle and lysogenic cycle (see Figure 10.3).
- Emerging viruses have arisen recently or are more likely to cause infection (see Figure 10.4).
- The disease AIDS is caused by a virus called human immunodeficiency virus (HIV). The virus is a retrovirus whose reproductive cycle involves the integration of the viral genome into a chromosome in the host cell (see Figure 10.5).

10.3 Plaque Formation and Intergenic Complementation in Bacteriophages

- Bacteriophages are viruses that infect bacteria (see Figure 10.6).
- Some bacteriophages cause bacteria to lyse. This event can lead to plaque formation when bacteria are plated as a lawn on bacterial media (see Figure 10.7).
- Studying a type of bacteriophage called T4, Benzer identified mutant strains that caused rapid lysis, thereby leading to larger plaques (see Figure 10.8).
- Benzer identified strains of T4 that could not form plaques in *E. coli* K12(λ). He conducted complementation tests to determine if the mutations in these strains were in the same phage gene or in different genes (see Figure 10.9).

10.4 Intragenic Mapping in Bacteriophages

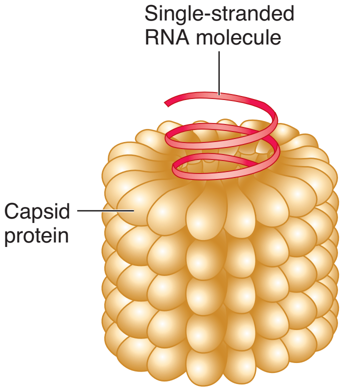
- Benzer devised a method to study intragenic recombination. He used deletion mapping to determine the locations of the mutations within two phage genes (see Figures 10.10–10.13).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Discuss how the components of viruses assemble to make new virus particles.

Answer: Some viruses with a simple structure self-assemble, meaning that viral components spontaneously bind to each other to form a complete virus particle. An example of a self-assembling virus is tobacco mosaic virus (TMV), which we examined earlier. The capsid proteins assemble around the RNA genome, which becomes trapped inside the hollow capsid.



Other viruses do not self-assemble. For example, the assembly of phage λ requires the help of noncapsid proteins not found in the completed phage particle. Some of these noncapsid proteins function as enzymes that modify capsid proteins, whereas others serve as scaffolding for the assembly of the capsid. The assembly of an HIV virus occurs in two stages. First, capsid proteins assemble around two molecules of HIV RNA and two molecules of reverse transcriptase. Next, the newly formed capsid acquires its outer envelope in a budding process. This second phase of assembly occurs as the virus is released from the cell.

S2. From the perspective of the bacteriophage, what is the advantage of having both a lytic and lysogenic cycle?

Answer: For phages such as λ that can follow either cycle, environmental conditions influence whether it follows the lytic or lysogenic cycle. If nutrients are readily available, phage λ usually proceeds directly to the lytic cycle after its DNA enters the cell. This allows the phage to rapidly proliferate. Alternatively, if nutrients are in short supply, the lysogenic cycle is usually favored because insufficient nutrients may be available to make new viruses. If more nutrients become available later, the prophage may become activated, and the viral reproductive cycle will switch to the lytic cycle. Therefore, under more favorable conditions, new viruses are made and released.

S3. What are three different mechanisms of viral latency?

Answer: Some bacteriophages enter the lysogenic cycle in which the phage DNA becomes incorporated into the bacterial chromosome as a prophage. Eukaryotic viruses can become latent in other ways. Retroviruses, which have an RNA genome, are reverse transcribed and the viral DNA is incorporated into a chromosome as a provirus. Alternatively, herpesviruses can exist as episomes that remain latent for long periods of time.

S4. What is an emerging virus? Give three examples. How do you think that emerging viruses arise?

Answer: Emerging viruses are viruses that have arisen recently or are more likely to cause infection. Examples include HIV, the corona virus that causes SARS, and the influenza strain called H1N1. Emerging viruses usually arise by new mutations. Such mutations may alter the ability of a virus to infect its host or alter its ability to cause disease symptoms.

Conceptual Questions

C1. Discuss why viruses are considered nonliving.

C2. What structural features are common to all viruses? Which features are found only in certain types of viruses?

C3. What are the similarities and differences among viral genomes?

C4. What is a viral envelope? Describe how it is made.

C5. What do the terms *host cell* and *host range* mean?

C6. Describe why the attachment step in a viral reproductive cycle is usually specific for one or just a few cell types.

C7. Compare and contrast the entry step of the viral reproductive cycle of phage λ and HIV.

C8. What is the role of reverse transcriptase in the reproductive cycle of HIV and other retroviruses?

C9. Describe how lytic bacteriophages are released from their host cells.

C10. What is the difference between a temperate phage versus a virulent phage?

C11. What is a prophage, a provirus, and an episome? What is their common role in a viral reproductive cycle?

C12. What does the term *complementation mean*? If two different mutations that produce the same phenotype can complement each other, what can you conclude about the locations of each mutation?

C13. Intragenic mapping is sometimes called interallelic mapping. Explain why the two terms mean the same thing. In your own words, explain what an intragenic map is.

C14. As discussed in Chapter 14, genes are composed of a sequence of nucleotides. A typical gene in a bacteriophage is a few hundred or a few thousand nucleotides in length. If two different strains of bacteriophage T4 have a mutation in the *rIIA* gene that gives a rapid-lysis phenotype, yet they never produce wild-type phages by intragenic recombination when they are coinfected into *E. coli* B, what would you conclude about the locations of the mutations in the two different T4 strains?

Application and Experimental Questions

E1. Discuss how researchers determined that TMV is a virus that causes damage to plants.

E2. What technique must be used to visualize a virus?

E3. Certain drugs to combat human viral diseases affect spike glycoproteins in the viral envelope. Discuss how you think that such drugs may prevent viral infection.

E4. If you analyzed a phage plaque from a petri plate, what would it contain?

E5. As shown in Figure 10.9, phages with *rII* mutations cannot produce plaques in *E. coli* K12(λ), but wild-type phages can. From an experimental point of view, explain why this observation is so significant.

E6. In the experimental strategy described in Figure 10.11, explain why it was necessary to dilute the phage preparation used to infect *E. coli* B so much more than the phage preparation used to infect *E. coli* K12(λ).

E7. Here are data from several complementation experiments, involving rapid-lysis mutations in genes *rIIA* and *rIIB*. The strain designated *L51* is known to have a mutation in *rIIB*.

Phage Mixture	Complementation
L91 and L65	No
L65 and L62	No
L33 and L47	Yes
L40 and L51	No
L47 and L92	No
L51 and L47	Yes
L51 and L92	Yes
L33 and L40	No
L91 and L92	Yes
L91 and L33	No

List which groups of mutations are in the *rIIA* gene and which groups are in the *rIIB* gene.

E8. A researcher has several different strains of T4 phage with single mutations in the same gene. In these strains, the mutations render the phage temperature-sensitive. This means that temperature-sensitive phages can propagate when the bacterium (*E. coli*) is grown at 32°C but cannot propagate themselves when *E. coli* is grown at 37°C. Think about Benzer's strategy for intragenic mapping and propose an experimental strategy to map the temperature-sensitive mutations.

E9. Explain how Benzer's results indicated that a gene is not an indivisible unit.

E10. Explain why deletion mapping was used as a step in the intragenic mapping of *rII* mutations.

Questions for Student Discussion/Collaboration

1. Discuss the advantages of the genetic analysis of viruses. What types of allelic differences among viruses are suitable for genetic analyses?

2. Complementation occurs when two defective alleles in two different genes are found within the same organism and produce a normal phenotype. What other examples of complementation have we encountered in previous chapters of this textbook?

Answers to Comprehension Questions

10.1: c, d

10.2: c, d

10.3: a, b

10.4: b, c

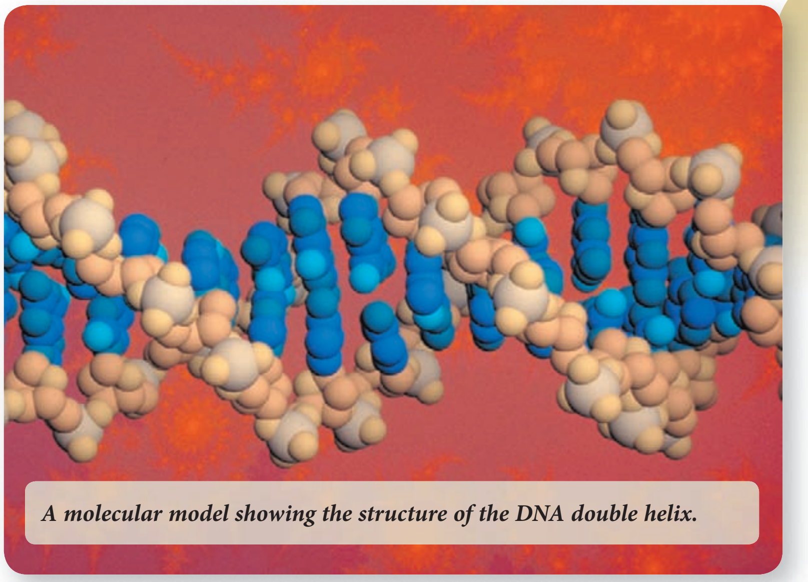
Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

PART III MOLECULAR STRUCTURE AND REPLICATION OF THE GENETIC MATERIAL

11


A molecular model showing the structure of the DNA double helix.

CHAPTER OUTLINE

11.1 Identification of DNA as the Genetic Material
11.2 Overview of DNA and RNA Structure
11.3 Nucleotide Structure
11.4 Structure of a DNA Strand
11.5 Discovery of the Double Helix
11.6 Structure of the DNA Double Helix
11.7 RNA Structure

MOLECULAR STRUCTURE OF DNA AND RNA

In Chapters 2 through 10, we focused on the relationship between the inheritance of genes and chromosomes and the outcome of an organism's traits. In Chapter 11, we will shift our attention to **molecular genetics**—the study of DNA structure and function at the molecular level. An exciting goal of molecular genetics is to use our knowledge of DNA structure to understand how DNA functions as the genetic material. Using molecular techniques, researchers have determined the organization of many genes. This information, in turn, has helped us understand how the expression of such genes governs the outcome of an individual's inherited traits.

The past several decades have seen dramatic advances in techniques and approaches to investigate and even to alter the genetic material. These advances have greatly expanded our understanding of molecular genetics and also have provided key insights into the mechanisms underlying transmission and population genetics. Molecular genetic technology is also widely used in supporting disciplines such as biochemistry, cell biology, and microbiology.

To a large extent, our understanding of genetics comes from our knowledge of the molecular structure of **DNA (deoxyribonucleic acid)** and **RNA (ribonucleic acid).** In this chapter, we will begin by considering classic experiments that showed that DNA is the genetic material. We will then survey the molecular features of DNA and RNA that underlie their function.

11.1 IDENTIFICATION OF DNA AS THE GENETIC MATERIAL

To fulfill its role, the genetic material must meet four criteria:

1. **Information:** The genetic material must contain the information necessary to construct an entire organism. In other words, it must provide the blueprint to determine the inherited traits of an organism.
2. **Transmission:** During reproduction, the genetic material must be passed from parents to offspring.
3. **Replication:** Because the genetic material is passed from parents to offspring, and from mother cell to daughter cells during cell division, it must be copied.
4. **Variation:** Within any species, a significant amount of phenotypic variability occurs. For example, Mendel studied several characteristics in pea plants that varied among different plants. These included height (tall versus dwarf)

and seed color (yellow versus green). Therefore, the genetic material must also vary in ways that can account for the known phenotypic differences within each species.

Recall from Chapter 3 that in the 1880s, August Weismann and Carl Nägeli championed the idea that a chemical substance within living cells is responsible for the transmission of traits from parents to offspring and that subsequent experimentation demonstrated that chromosomes are the carriers of the genetic material. Nevertheless, the story was not complete because chromosomes contain both DNA and proteins. Also, RNA is found in the vicinity of chromosomes. Therefore, further research was needed to precisely identify the genetic material. In this section, we will examine the first experimental approaches to achieve this goal.

Griffith's Experiments Indicated That Genetic Material Can Transform Pneumococcus

Some early work in microbiology was important in developing an experimental strategy to identify the genetic material. Frederick Griffith studied a type of bacterium known then as pneumococci and now classified as *Streptococcus pneumoniae*. Certain strains of *S. pneumoniae* secrete a polysaccharide capsule, whereas other strains do not. When streaked onto petri plates containing a solid growth medium, capsule-secreting strains have a smooth colony morphology, whereas those strains unable to secrete a capsule have a rough appearance.

The different forms of *S. pneumoniae* also affect their virulence, or ability to cause disease. When smooth strains of *S. pneumonia* infect a mouse, the capsule allows the bacteria to escape attack by the mouse's immune system. As a result, the bacteria can grow and eventually kill the mouse. In contrast, the nonencapsulated (rough) bacteria are destroyed by the animal's immune system.

In 1928, Griffith conducted experiments that involved the injection of live and/or heat-killed bacteria into mice. He then observed whether or not the bacteria caused a lethal infection. Griffith was working with two strains of *S. pneumoniae*, a type S (S for smooth) and a type R (R for rough). When injected into a live mouse, the type S bacteria proliferated within the mouse's bloodstream and ultimately killed the mouse (**Figure 11.1a**). Following the death of the mouse, Griffith found many type S bacteria within the mouse's blood. In contrast, when type R bacteria were injected into a mouse, the mouse lived (**Figure 11.1b**). To verify that the proliferation of the smooth bacteria was causing the death of the mouse, Griffith killed the smooth bacteria with heat treatment before injecting them into the mouse. In this case, the mouse also survived (**Figure 11.1c**).

The critical and unexpected result was obtained in the experiment outlined in **Figure 11.1d**. In this experiment, live type R bacteria were mixed with heat-killed type S bacteria. As shown here, the mouse died. Furthermore, extracts from tissues of the dead mouse were found to contain living type S bacteria! What can account for these results? Because living type R bacteria alone could not proliferate and kill the mouse (Figure 11.1b), the interpretation of the data in Figure 11.1d is that something from the dead type S bacteria was transforming the type R bacteria into type S. Griffith called this process **transformation,** and

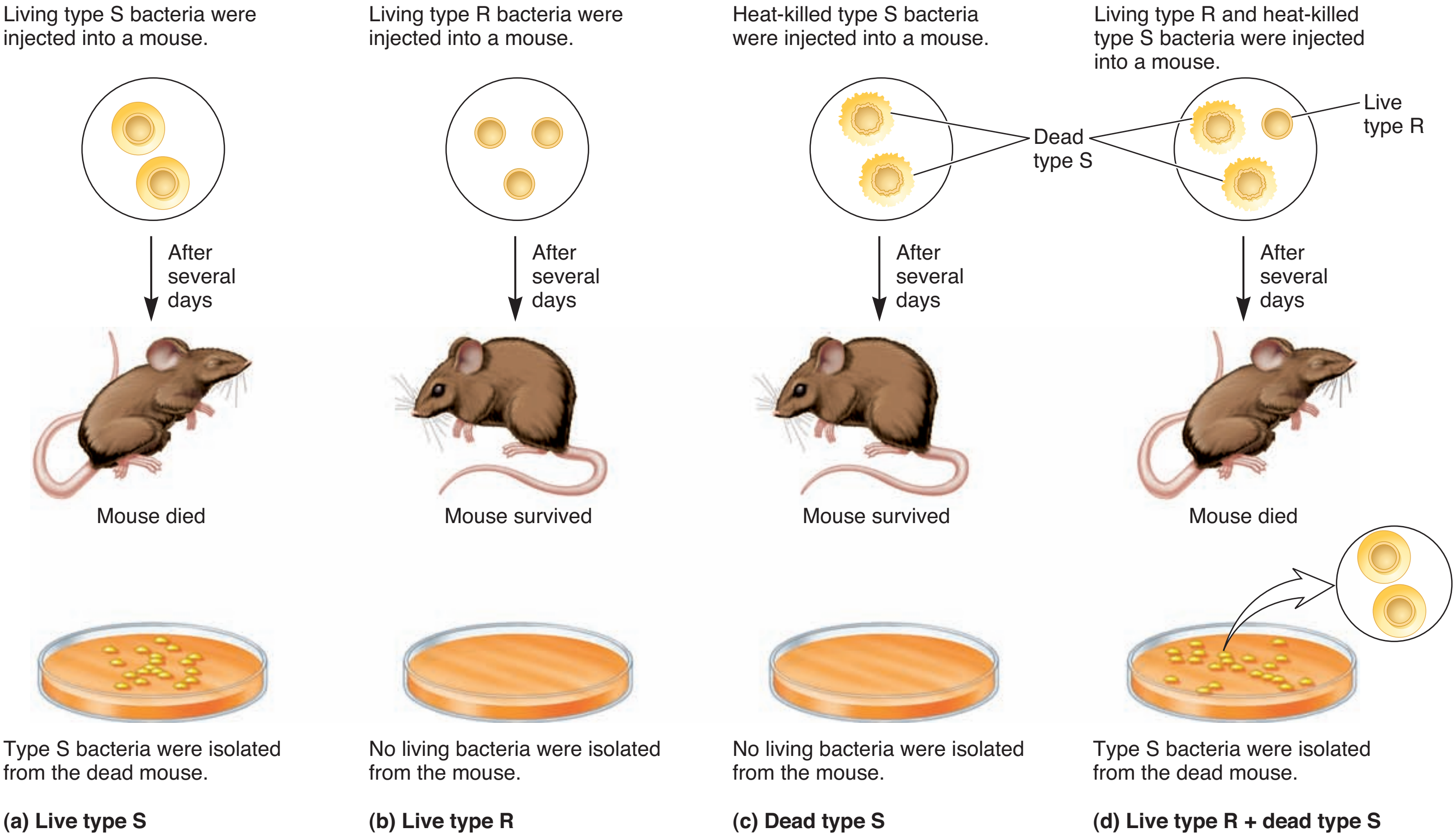

Living type S bacteria were injected into a mouse.
Living type R bacteria were injected into a mouse.
Heat-killed type S bacteria were injected into a mouse.
Living type R and heat-killed type S bacteria were injected into a mouse.
Dead type S
Live type R
After several days
After several days
After several days
After several days
Mouse died
Mouse survived
Mouse survived
Mouse died
Type S bacteria were isolated from the dead mouse.
No living bacteria were isolated from the mouse.
No living bacteria were isolated from the mouse.
Type S bacteria were isolated from the dead mouse.
(a) Live type S
(b) Live type R
(c) Dead type S
(d) Live type R + dead type S

FIGURE 11.1 **Griffith's experiments on genetic transformation in pneumococcus.**

Concept check: *Explain why the mouse in part (d) died.*

the unidentified substance causing this to occur was termed the transforming principle. The steps of bacterial transformation are described in Chapter 9 (see Figure 9.12).

At this point, let's look at what Griffith's observations mean in terms of the criteria for the genetic material. The transformed bacteria acquired the *information* to make a capsule. Among different strains, *variation* exists in the ability to create a capsule and to cause mortality in mice. The genetic material that is necessary to create a capsule must be *replicated* so that it can be *transmitted* from mother to daughter cells during cell division. Taken together, these observations are consistent with the idea that the formation of a capsule is governed by the bacteria's genetic material, meeting the four criteria described previously. Griffith's experiments showed that some genetic material from the dead bacteria was transferred to the living bacteria and provided them with a new trait. However, Griffith did not know what the transforming substance was.

Avery, MacLeod, and McCarty Showed that DNA Is the Substance That Transforms Bacteria

Important scientific discoveries often take place when researchers recognize that someone else's experimental observations can be used to address a particular scientific question. Oswald Avery, Colin MacLeod, and Maclyn McCarty realized that Griffith's observations could be used as part of an experimental strategy to identify the genetic material. They asked the question, "What substance is being transferred from the dead type S bacteria to the live type R?" To answer this question, they incorporated additional biochemical techniques into their experimental methods.

At the time of these experiments in the 1940s, researchers already knew that DNA, RNA, proteins, and carbohydrates are major constituents of living cells. To separate these components and to determine if any of them was the genetic material, Avery, MacLeod, and McCarty used established biochemical purification procedures and prepared bacterial extracts from type S strains containing each type of these molecules. After many repeated attempts with different types of extracts, they discovered that only one of the extracts, namely, the one that contained purified DNA, was able to convert the type R bacteria into type S. As shown in **Figure 11.2**, when this extract was mixed with type R bacteria, some of the bacteria were converted to type S. However, if no DNA extract was added, no type S bacterial colonies were observed on the petri plates.

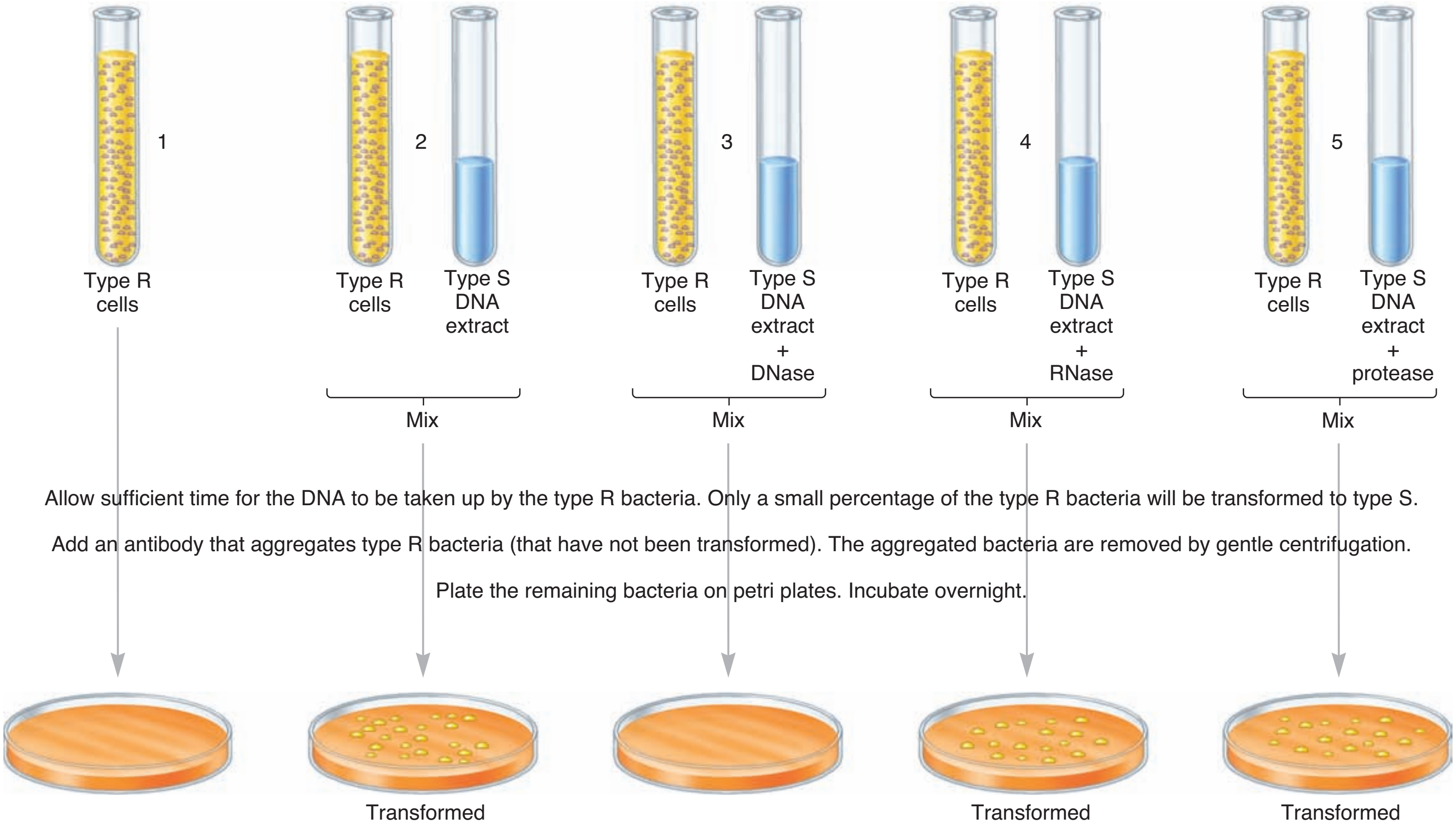

1
Type R cells
2
Type R cells
Type S DNA extract
Mix
3
Type R cells
Type S DNA extract + DNase
Mix
4
Type R cells
Type S DNA extract + RNase
Mix
5
Type R cells
Type S DNA extract + protease
Mix
Allow sufficient time for the DNA to be taken up by the type R bacteria. Only a small percentage of the type R bacteria will be transformed to type S.
Add an antibody that aggregates type R bacteria (that have not been transformed). The aggregated bacteria are removed by gentle centrifugation.
Plate the remaining bacteria on petri plates. Incubate overnight.
Transformed
Transformed
Transformed

FIGURE 11.2 **Experimental protocol used by Avery, MacLeod, and McCarty to identify the transforming principle.** Samples of *Streptococcus pneumoniae* cells were either not exposed to a type S DNA extract (tube 1) or exposed to a type S DNA extract (tubes 2–5). Tubes 3, 4, and 5 also contained DNase, RNase, or protease, respectively. After incubation, the cells were exposed to antibodies, which are molecules that can specifically recognize the molecular structure of macromolecules. In this experiment, the antibodies recognized the cell surface of type R bacteria and caused them to clump together. The clumped bacteria were removed by a gentle centrifugation step. Only the bacteria that were not recognized by the antibody (namely, the type S bacteria) remained in the supernatant. The cells in the supernatant were plated on solid growth media. After overnight incubation, visible colonies may be observed.

Concept check: *What was the purpose of adding RNase and protease to a DNA extract?*

A biochemist might point out that a DNA extract may not be 100% pure. In fact, any purified extract might contain small traces of some other substances. Therefore, one can argue that a small amount of contaminating material in the DNA extract might actually be the genetic material. The most likely contaminating substances in this case would be RNA or protein. To further verify that the DNA in the extract was indeed responsible for the transformation, Avery, MacLeod, and McCarty treated samples of the DNA extract with enzymes that digest DNA (called **DNase**), RNA **(RNase)**, or protein **(protease)** (see Figure 11.2). When the DNA extracts were treated with RNase or protease, they still converted type R bacteria into type S. These results indicated that any remaining RNA or protein in the extract was not acting as the genetic material. However, when the extract was treated with DNase, it lost its ability to convert type R into type S bacteria. These results indicated that the degradation of the DNA in the extract by DNase prevented conversion of type R to type S. This interpretation is consistent with the hypothesis that DNA is the genetic material. A more elegant way of saying this is that the transforming principle is DNA.

Hershey and Chase Provided Evidence That DNA Is the Genetic Material of T2 Phage

A second experimental approach indicating that DNA is the genetic material came from the studies of Alfred Hershey and Martha Chase in 1952. Their research centered on the study of a virus known as T2. This virus infects *Escherichia coli* bacterial cells and is therefore known as a **bacteriophage,** or simply a **phage.** As shown in **Figure 11.3**, the external structure of the T2 phage,

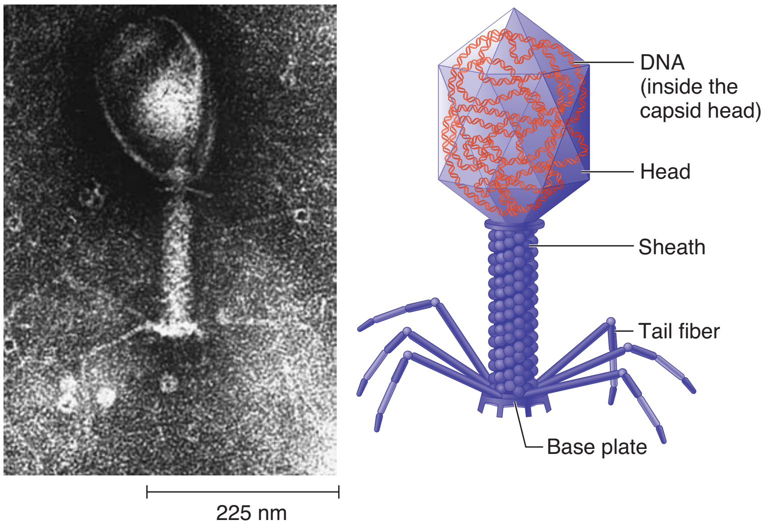
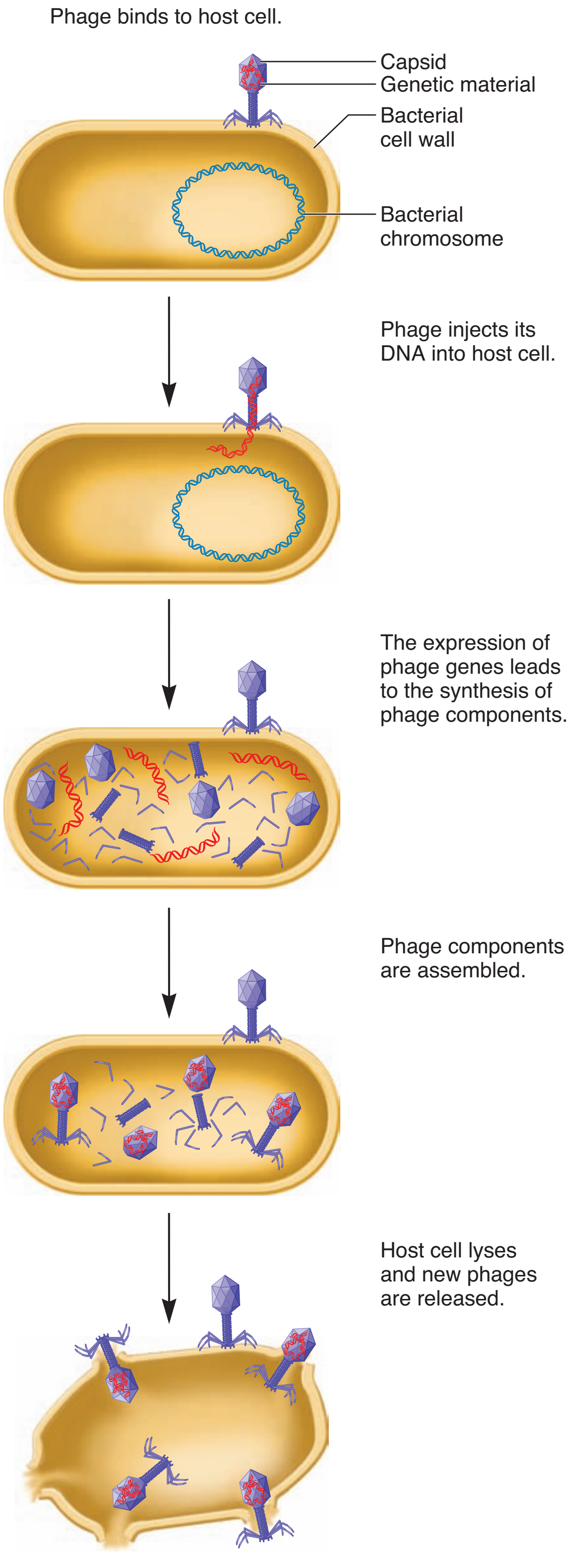

DNA (inside the capsid head)
Head
Sheath
Tail fiber
Base plate
225 nm

FIGURE 11.3 **Structure of the T2 bacteriophage.** The T2 bacteriophage is composed of a phage coat, or capsid, with genetic material inside the head of the capsid. The capsid is divided into regions called the head, sheath, tail fibers, and base plate. These components are composed of proteins. The genetic material is composed of DNA.

Genes → Traits The genetic material of a bacteriophage contains many genes, which provide the blueprint for making new viruses. When the bacteriophage injects its genetic material into a bacterium, these genes are activated and direct the host cell to make new bacteriophages, as described in Figure 11.4.


Phage binds to host cell.
Capsid
Genetic material
Bacterial cell wall
Bacterial chromosome
Phage injects its DNA into host cell.
The expression of phage genes leads to the synthesis of phage components.
Phage components are assembled.
Host cell lyses and new phages are released.

ONLINE ANIMATION

FIGURE 11.4 **Reproductive cycle of the T2 bacteriophage.**

Concept check: For infection to occur, what has to happen to the T2 genetic material?

known as the capsid or phage coat, consists of a head, sheath, tail fibers, and base plate. Biochemically, the phage coat is composed entirely of protein, which includes several different polypeptides. DNA is found inside the head of the T2 capsid. From a molecular point of view, this virus is rather simple, because it is composed of only two types of macromolecules: DNA and proteins.

Although the viral genetic material contains the blueprint to make new viruses, a virus itself cannot synthesize new viruses. As discussed in Chapter 10, a virus must introduce its genetic material into the cytoplasm of a living cell. In the case of T2, this first involves the attachment of its tail fibers to the bacterial cell wall and the subsequent injection of its genetic material into the cytoplasm of the cell (**Figure 11.4**). The phage coat remains attached on the outside of the bacterium and does not enter the cell. After the entry of the viral genetic material, the bacterial cytoplasm provides all of the machinery necessary to make viral proteins and DNA. The viral proteins and DNA assemble to make new viruses that are subsequently released from the cell by **lysis** (i.e., cell breakage).

To verify that DNA is the genetic material of T2, Hershey and Chase devised a method for separating the phage coat, which is attached to the outside of the bacterium, from the genetic material, which is injected into the cytoplasm. They were aware of microscopy experiments by Thomas Anderson showing that the T2 phage attaches itself to the outside of a bacterium by its tail fibers. Hershey and Chase reasoned that this is a fairly precarious attachment that could be disrupted by subjecting the bacteria to high shear forces, such as those produced in a kitchen blender. Their method was to expose bacteria to T2 phage, allowing sufficient time for the viruses to attach to bacteria and inject their genetic material. They then sheared the phage coats from the surface of the bacteria by a blender treatment. In this way, the phages' genetic material, which had been injected into the cytoplasm of the bacterial cells, could be separated from the phage coats that were sheared away.

Hershey and Chase used radioisotopes to distinguish proteins from DNA. Sulfur atoms are found in proteins but not in DNA, whereas phosphorus atoms are found in DNA but not in phage proteins. Therefore, ^{35}S (a radioisotope of sulfur) and ^{32}P (a radioisotope of phosphorus) were used to specifically label phage proteins and DNA, respectively. In step 1 of the experiment described in **Figure 11.5**, the researchers grew *E. coli* cells in media that contained ^{35}S or ^{32}P and then infected the *E. coli* cells with T2 phages. When new phages were produced, they were labeled with ^{35}S or ^{32}P. One preparation was labeled with ^{35}S to label the phage proteins, and the other preparation was labeled with ^{32}P to label the phage DNA.

Next, each type of labeled phage was mixed with a new sample of *E. coli* cells (see step 2). The phages were given sufficient time to inject their genetic material into the bacterial cells, and then the samples were subjected to shearing force using a blender. This treatment was expected to remove the phage coats from the surfaces of the bacterial cells. The samples were then subjected to centrifugation at a speed that would cause the heavier bacterial cells to form a pellet at the bottom of the tube, whereas the light phage coats would remain in the supernatant, the liquid found above the pellet. The amount of radioactivity in the supernatant (emitted from either ^{35}S or ^{32}P) was determined using a Geiger counter.

▶ THE HYPOTHESIS

Only the genetic material of the phage is injected into the bacterium. Isotope labeling will reveal if it is DNA or protein.

▶ TESTING THE HYPOTHESIS — FIGURE 11.5 — Evidence that DNA is the genetic material of T2 bacteriophage.

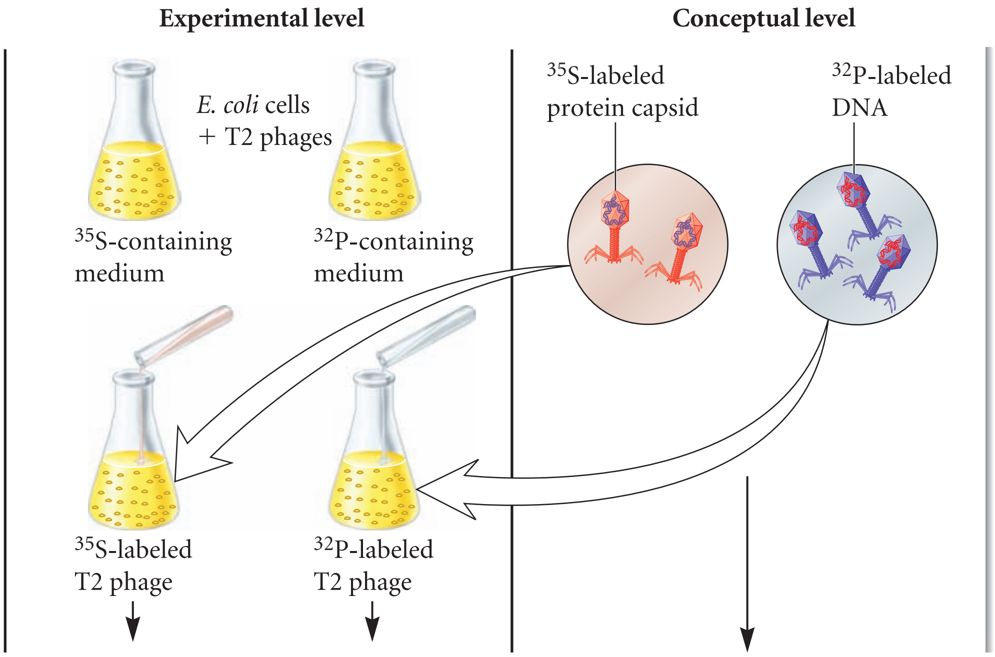
Starting materials: The starting materials were *E. coli* cells and T2 phage.

Concept check: *Why were two different radioisotopes used?*

1. Grow *E. coli* cells in media containing ^{35}S or ^{32}P. Infect with T2 phage. This will produce new T2 phages that are labeled with ^{35}S or ^{32}P.

2. Into one flask, add ^{35}S-labeled phage; in the second flask, add ^{32}P-labeled phage. Add uninfected *E. coli* cells. Note: the media does not contain ^{35}S or ^{32}P.

3. Allow infection to occur.

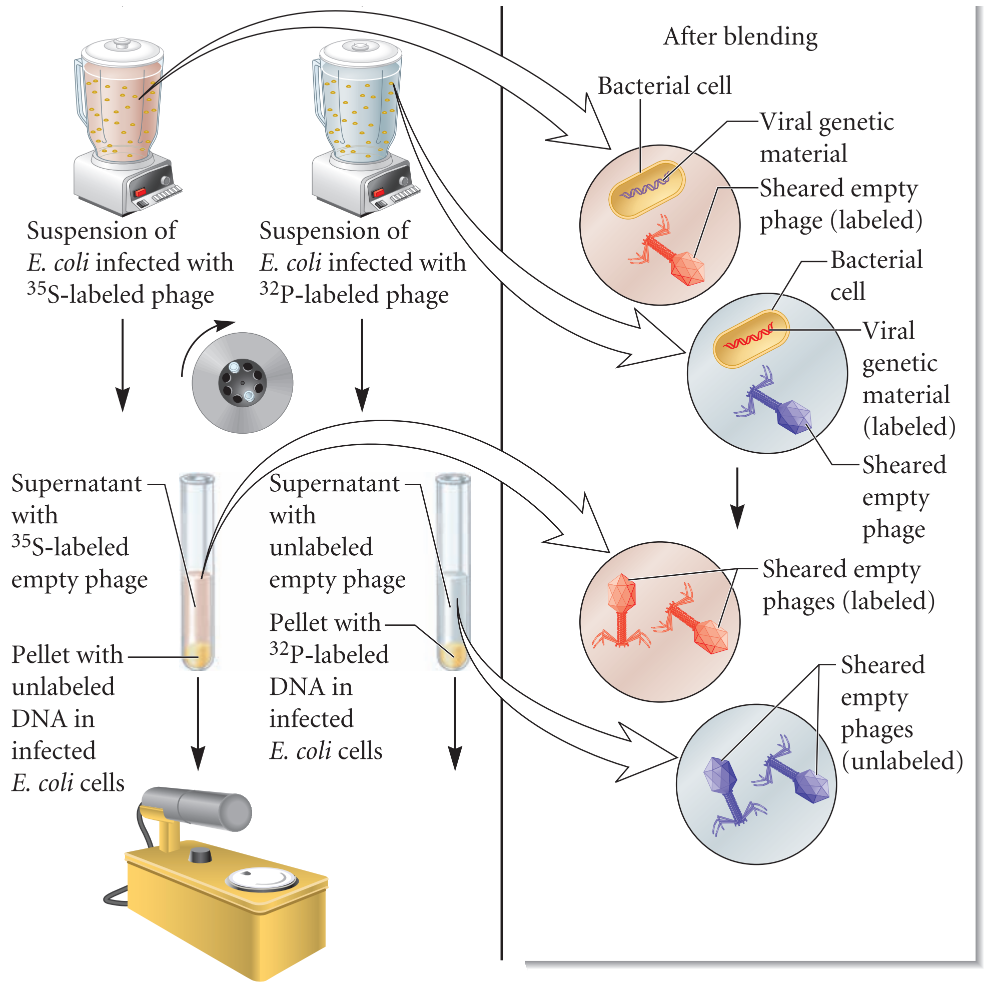


4. Agitate solutions in blenders for different lengths of time to shear the empty phages off the bacterial cells.

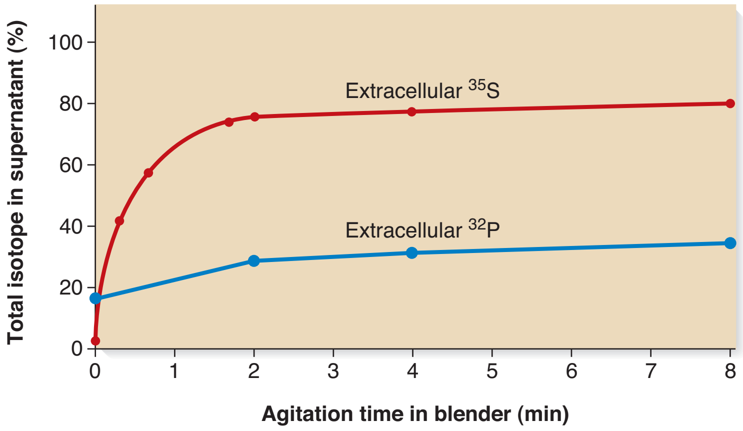
5. Centrifuge at 10,000 rpm.

6. The heavy bacterial cells sediment to the pellet, while the lighter phages remain in the supernatant. (See Appendix for explanation of centrifugation.)

7. Count the amount of radioisotope in the supernatant with a Geiger counter. Compare it with the starting amount.


After blending
Suspension of E. coli infected with ^{35}S-labeled phage
Suspension of E. coli infected with ^{32}P-labeled phage
Bacterial cell
Viral genetic material
Sheared empty phage (labeled)
Bacterial cell
Viral genetic material (labeled)
Sheared empty phage
Supernatant with ^{35}S-labeled empty phage
Pellet with unlabeled DNA in infected E. coli cells
Supernatant with unlabeled empty phage
Pellet with ^{32}P-labeled DNA in infected E. coli cells
Sheared empty phages (labeled)
Sheared empty phages (unlabeled)

THE DATA



80%
Blending removes 80% of ^{35}S from *E. coli* cells.

35%
Most of the ^{32}P (65%) remains with intact *E. coli* cells.

Data from A. D. Hershey and Martha Chase (1952) Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *Journal of General Physiology* 36, 39–56.

INTERPRETING THE DATA

As seen in the data, most of the ^{35}S isotope was found in the supernatant. Because the shearing force was expected to remove the phage coat, this result indicates that the empty phages contain primarily protein. By comparison, only about 35% of the ^{32}P was found in the supernatant following shearing. Therefore, most of the DNA was located within the bacterial cells in the pellet. These results are consistent with the idea that the DNA is injected into the bacterial cytoplasm during infection, which would be the expected result if DNA is the genetic material.

By themselves, the results described in Figure 11.5 were not conclusive evidence that DNA is the genetic material. For example, you may have noticed that less than 100% of the phage protein was found in the supernatant. Therefore, some of the phage protein could have been introduced into the bacterial cells (and could function as the genetic material). Nevertheless, the results of Hershey and Chase were consistent with the conclusion that the genetic material is DNA rather than protein. Overall, their studies of T2 phage were quite influential in convincing the scientific community that DNA is the genetic material.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

RNA Functions as the Genetic Material in Some Viruses

We now know that bacteria, archaea, protists, fungi, plants, and animals all use DNA as their genetic material. As mentioned, viruses also have their own genetic material. Hershey and Chase concluded from their experiments that this genetic material is DNA. In the case of T2 bacteriophage, that is the correct conclusion. However, as described in Chapter 10, many viruses use RNA, rather than DNA, as their genetic material. In 1956, Alfred Gierer and Gerhard Schramm isolated RNA from the tobacco mosaic virus (TMV), which infects plant cells. When this purified RNA was applied to plant tissue, the plants developed the same types of lesions that occurred when they were exposed to intact tobacco mosaic viruses. Gierer and Schramm correctly concluded that the viral genome of TMV is composed of RNA. As discussed in Chapter 10, many other viruses have been found to contain RNA as their genetic material (refer back to Table 10.1).

11.1 REVIEWING THE KEY CONCEPTS

- To fulfill its role, a genetic material must meet four criteria: information, transmission, replication, and variation.
- Griffith showed that the genetic material from type S bacteria could transform type R bacteria into type S (see Figure 11.1).
- Avery, MacLeod, and McCarty discovered that the transforming substance is DNA (see Figure 11.2).
- Hershey and Chase determined that the genetic material of T2 phage is DNA (see Figures 11.3–11.5).
- Viruses may use DNA or RNA as their genetic material.

11.1 COMPREHENSION QUESTIONS

1. In the experiment of Avery, MacLeod, and McCarty, the addition of RNase and protease to the DNA extracts
    a. prevented the conversion of type S bacteria into type R bacteria.
    b. allowed the conversion of type S bacteria into type R bacteria.
    c. prevented the conversion of type R bacteria into type S bacteria.
    d. allowed the conversion of type R bacteria into type S bacteria.
2. In the Hershey and Chase experiment involving T2 phage, after blender treatment and centrifugation
    a. most of the ^{32}P was in the supernatant and most of the ^{35}S was in the pellet.
    b. most of the ^{35}S was in the supernatant and most of the ^{32}P was in the pellet.
    c. equal amounts of ^{32}P and ^{35}S were in the supernatant and pellet.
    d. none of the above was observed.

11.2 OVERVIEW OF DNA AND RNA STRUCTURE

DNA and its molecular cousin, RNA, are known as **nucleic acids.** This term is derived from the discovery of DNA by Friedrich Miescher in 1869. He identified a novel phosphorus-containing substance from the nuclei of white blood cells found in waste surgical bandages. He named this substance nuclein. As the structure of DNA and RNA became better understood, it was found they are acidic molecules, which means they release hydrogen ions (H^+) in solution and have a net negative charge at neutral pH. Thus, the name *nucleic acid* was coined.

Both DNA and RNA are macromolecules composed of smaller building blocks. To fully appreciate their structures, we need to consider four levels of complexity (Figure 11.6):

1. **Nucleotides** form the repeating structural unit of nucleic acids.
2. Nucleotides are linked together in a linear manner to form a **strand** of DNA or RNA.
3. Two strands of DNA (and sometimes RNA) interact with each other to form a **double helix.**
4. The three-dimensional structure of DNA results from the folding and bending of the double helix. Within living cells, DNA is associated with a wide variety of proteins that influence its structure. Chapter 12 examines the roles of these proteins in creating the three-dimensional structure of DNA found within chromosomes.

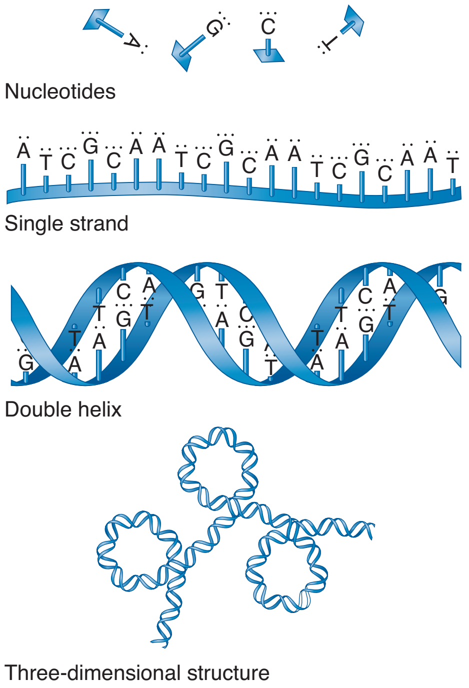


FIGURE 11.6 **Levels of nucleic acid structure.**

11.2 REVIEWING THE KEY CONCEPTS

- DNA and RNA are types of nucleic acids.
- In DNA, nucleotides are linked together to form strands, which then form a double helix that is found within chromosomes (see Figure 11.6).

11.2 COMPREHENSION QUESTIONS

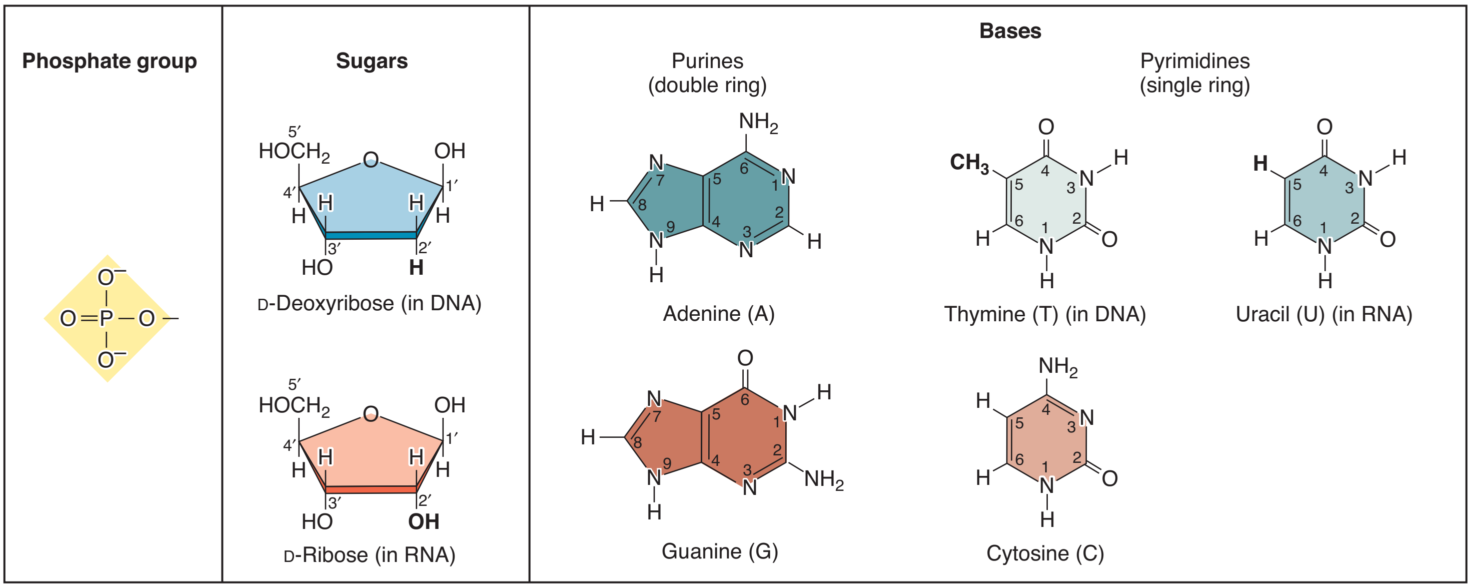
1. When going from simple to complex, which of the following is the proper order for DNA?
   a. Nucleotide, double helix, DNA strand, chromosome
   b. Nucleotide, chromosome, double helix, DNA strand
   c. Nucleotide, DNA strand, double helix, chromosome
   d. Chromosome, nucleotide, DNA strand, double helix

11.3 NUCLEOTIDE STRUCTURE

The nucleotide is the repeating structural unit of both DNA and RNA. A nucleotide has three components: at least one phosphate group, a pentose sugar, and a nitrogenous base. As shown in Figure 11.7, nucleotides can vary with regard to the sugar and the nitrogenous base. The two types of sugars are **deoxyribose** and **ribose,** which are found in DNA and RNA, respectively. The five different bases are subdivided into two categories: the **purines** and the **pyrimidines.** The purine bases, **adenine (A)** and **guanine (G),** contain a double-ring structure; the pyrimidine bases, **thymine (T), cytosine (C),** and **uracil (U),** contain a single-ring structure. Also, the base thymine is not found in RNA. Rather, uracil is found in RNA instead of thymine. Adenine, guanine, and cytosine occur in both DNA and RNA. As shown in Figure 11.7, the bases and sugars have a standard numbering system. The nitrogen and carbon atoms found in the ring structure of the bases are given numbers 1 through 9 for the purines and 1 through 6 for the pyrimidines. In comparison, the five carbons found in the sugars are designated with primes, such as 1′, to distinguish them from the numbers found in the bases.

Figure 11.8 shows the repeating unit of nucleotides found in DNA and RNA. The locations of the attachment sites of the base and phosphate to the sugar molecule are important to the nucleotide's function. In the sugar ring, carbon atoms are numbered in a clockwise direction, beginning with a carbon atom adjacent to the ring oxygen atom. The fifth carbon is outside the ring structure. In a single nucleotide, the base is always attached to the 1′ carbon atom, and one or more phosphate groups are attached at the 5′ position. As discussed later, the —OH group attached to the 3′ carbon is important in allowing nucleotides to form covalent linkages with each other.

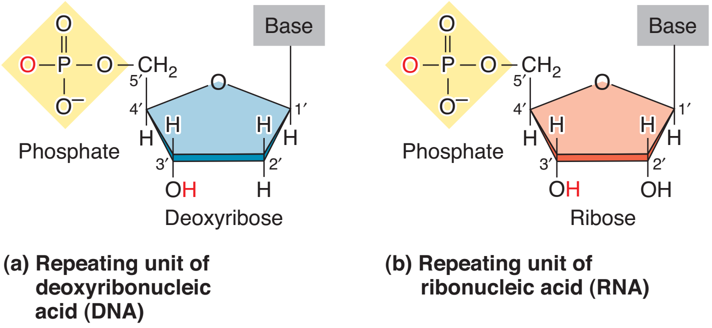
The terminology used to describe nucleic acid units is based on three structural features: the type of base, the type of sugar, and the number of phosphate groups. When a base is attached to only a sugar, we call this pair a **nucleoside.** If adenine is attached to ribose, this nucleoside is called adenosine (Figure 11.9). Nucleosides containing guanine, thymine, cytosine, or uracil are called guanosine, thymidine, cytidine, and uridine, respectively. When only the bases are attached to deoxyribose, they are called deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxycytidine. The covalent attachment of one or more phosphate molecules to a nucleoside creates a nucleotide. If a nucleotide contains adenine, ribose, and one phosphate, it is adenosine monophosphate, abbreviated AMP. If a nucleotide contains adenine, ribose, and three phosphate groups, it is called adenosine triphosphate, or ATP. If it contains guanine, ribose, and three phosphate groups, it is guanosine triphosphate, or GTP. A nucleotide composed of adenine, deoxyribose, and three phosphate groups is called deoxyadenosine triphosphate (dATP).



INTERACTIVE EXERCISE

FIGURE 11.7 **The components of nucleotides.** The three building blocks of a nucleotide are one or more phosphate groups, a sugar, and a base. The bases are categorized as purines (adenine and guanine) and pyrimidines (thymine, cytosine, and uracil).

Concept check: Which of these components are not found in DNA?

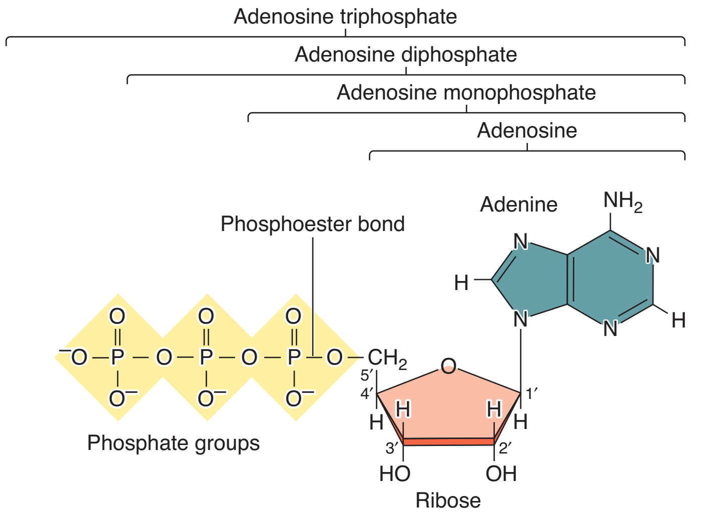

O
O–P–O–CH2
O–
Base
O
5′
4′
1′
H H
H H
3′
2′
OH H
Phosphate
Deoxyribose
O
O–P–O–CH2
O–
Base
O
5′
4′
1′
H H
H H
3′
2′
OH OH
Phosphate
Ribose

(a) Repeating unit of deoxyribonucleic acid (DNA)

(b) Repeating unit of ribonucleic acid (RNA)

INTERACTIVE EXERCISE

FIGURE 11.8 The structure of nucleotides found in (a) DNA and (b) RNA. DNA contains deoxyribose as its sugar and the bases A, T, G, and C. RNA contains ribose as its sugar and the bases A, U, G, and C. In a DNA or RNA strand, the oxygen on the 3′ carbon is linked to the phosphorus atom of phosphate in the adjacent nucleotide. The two atoms (O and H) shown in red are found within individual nucleotides but not when nucleotides are joined together to make strands of DNA and RNA.


Adenosine triphosphate
Adenosine diphosphate
Adenosine monophosphate
Adenosine
Phosphoester bond
Adenine
NH2
N
N
H
N
N
H
–O–P–O–P–O–P–O–CH2
O O O
O– O– O–
5′
4′
O
1′
H H
H H
3′
2′
HO OH
Phosphate groups
Ribose

INTERACTIVE EXERCISE

FIGURE 11.9 A comparison between the structures of an adenine-containing nucleoside and nucleotides. A nucleoside contains a sugar and base, whereas a nucleotide has a sugar, a base, and one or more phosphates.

11.3 REVIEWING THE KEY CONCEPTS

- A nucleotide is composed of one or more phosphates, a sugar, and a base. The purine bases are adenine and guanine, whereas the pyrimidine bases are thymine (DNA only), cytosine, and uracil (RNA only) The sugar deoxyribose is found in DNA, whereas ribose is in RNA (see Figures 11.7–11.8).
- A nucleoside contains a sugar and a base, whereas a nucleotide contains a sugar, a base, and one or more phosphate groups (see Figure 11.9).

11.3 COMPREHENSION QUESTIONS

1. Which of the following could be the components of a single nucleotide found in DNA?
   a. Deoxyribose, adenine, and thymine
   b. Ribose, phosphate, and cytosine
   c. Deoxyribose, phosphate, and thymine
   d. Ribose, phosphate, and uracil
2. A key difference between the nucleotides found in DNA versus RNA is that
   a. DNA has phosphate but RNA does not.
   b. DNA has deoxyribose, but RNA has ribose.
   c. DNA has thymine, but RNA has uracil.
   d. both b and c are correct.

11.4 STRUCTURE OF A DNA STRAND

A strand of DNA or RNA has many nucleotides that are covalently attached to each other in a linear fashion; DNA and RNA strands are **polynucleotides.** Figure 11.10 depicts a short strand of DNA with four nucleotides. A few structural features are worth noting. First, the linkage involves an ester bond between

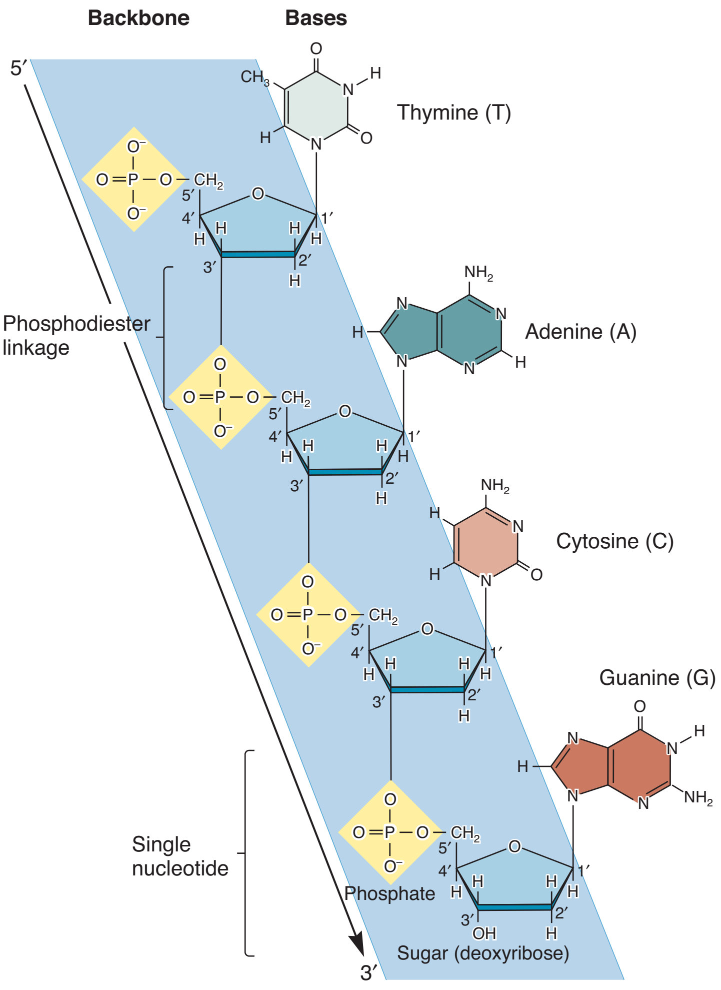

Backbone
Bases
5′
3′
Thymine (T)
Adenine (A)
Cytosine (C)
Guanine (G)
Phosphodiester linkage
Single nucleotide
Phosphate
Sugar (deoxyribose)
CH3
NH2
O=P–O–CH2
O–
5′
4′
3′
2′
1′
OH
H

FIGURE 11.10 A short strand of DNA containing four nucleotides. Nucleotides are covalently linked together to form a strand of DNA.

Concept check: Which components of a nucleotide form the backbone of a DNA strand?

a phosphate group on one nucleotide and the sugar molecule on the adjacent nucleotide. Another way of viewing this linkage is to notice that a phosphate group connects two sugar molecules. For this reason, the linkage in DNA or RNA strands is called a **phosphodiester linkage.** The phosphates and sugar molecules form the **backbone** of a DNA or RNA strand. The bases project from the backbone. The backbone is negatively charged due to a negative charge on each phosphate.

A second important structural feature is the orientation of the nucleotides. As mentioned, the carbon atoms in a sugar molecule are numbered in a particular way. A phosphodiester linkage involves a phosphate attachment to the 5′ carbon in one nucleotide and to the 3′ carbon in the other. In a strand, all sugar molecules are oriented in the same direction. As shown in Figure 11.10, the 5′ carbons in every sugar molecule are above the 3′ carbons. Therefore, a strand has a **directionality** based on the orientation of the sugar molecules within that strand. In Figure 11.10, the direction of the strand is 5′ to 3′ when going from top to bottom.

A critical aspect regarding DNA and RNA structure is that a strand contains a specific sequence of bases. In Figure 11.10, the sequence of bases is thymine–adenine–cytosine–guanine, abbreviated TACG. Furthermore, to show the directionality, the strand should be abbreviated 5′–TACG–3′. Because the nucleotides within a strand are covalently attached to each other, the sequence of bases cannot shuffle around and become rearranged. Therefore, the sequence of bases in a DNA strand remains the same over time, except in rare cases when mutations occur. As we will see throughout this textbook, the sequence of bases within DNA and RNA is the defining feature that allows them to carry information.

11.4 REVIEWING THE KEY CONCEPTS

- In a DNA strand, nucleotides are covalently attached to one another via phosphodiester linkages (see Figure 11.10).

11.4 COMPREHENSION QUESTIONS

1. In a DNA strand, a phosphate connects a 3′ carbon atom in one deoxyribose to
   a. a 5′ carbon in an adjacent deoxyribose.
   b. a 3′ carbon in an adjacent deoxyribose.
   c. a base in an adjacent nucleotide
   d. none of the above.

11.5 DISCOVERY OF THE DOUBLE HELIX

A major discovery in molecular genetics was made in 1953 by James Watson and Francis Crick. At that time, DNA was already known to be composed of nucleotides. However, it was not understood how the nucleotides are bonded together to form the structure of DNA. Watson and Crick committed themselves to determine the structure of DNA because they felt this knowledge was needed to understand the functioning of genes. Other researchers, such as Rosalind Franklin and Maurice Wilkins, shared this view. Before we examine the characteristics of the double helix, let's consider the events that provided the scientific framework for Watson and Crick's breakthrough.

A Few Key Events Led to the Discovery of the Double-Helix Structure

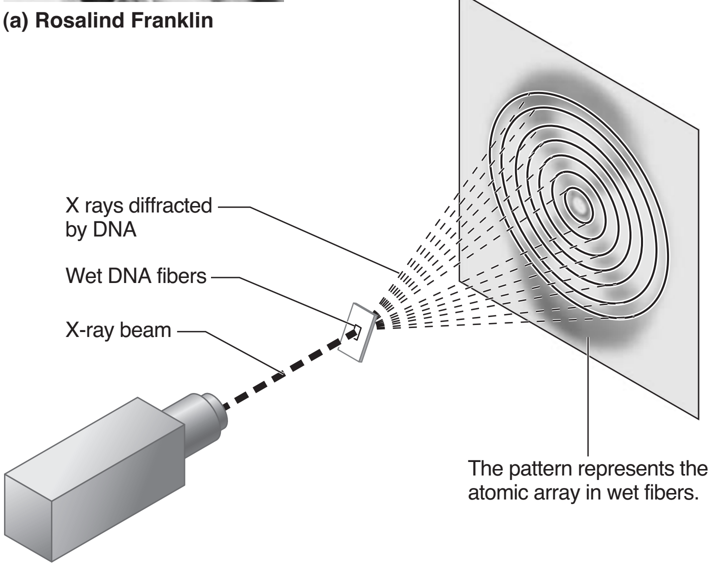
In the early 1950s, Linus Pauling proposed that regions of proteins can fold into a particular type of secondary structure. To elucidate this structure, Pauling built large models by linking together simple ball-and-stick units (**Figure 11.11**). By carefully scaling the objects in his models, he could visualize if atoms fit together properly in a complicated three-dimensional structure. Is this approach still used today? The answer is "Yes," except that today researchers construct their three-dimensional models on computers. As we will see, Watson and Crick also used a ball-and-stick approach to solve the structure of the DNA double helix. Interestingly, they were well aware that Pauling might figure out the structure of DNA before they did. This provided a stimulating rivalry between the researchers.

A second important development that led to the elucidation of the double helix was X-ray diffraction data. When a purified substance, such as DNA, is subjected to X-rays, it produces a well-defined diffraction pattern if the molecule is organized into a regular structural pattern. An interpretation of the diffraction pattern (using mathematical theory) can ultimately provide information concerning the structure of the molecule. Rosalind Franklin (**Figure 11.12a**), working in the same laboratory as Maurice Wilkins, used X-ray diffraction to study wet DNA fibers. Franklin made marked advances in X-ray diffraction techniques while working with DNA. She adjusted her equipment to produce an extremely fine beam of X-rays. She extracted finer DNA fibers than ever before and arranged them in parallel bundles.

FIGURE 11.11 **The use of models by Linus Pauling.**

Concept check: *Why is modeling useful?*

(a) Rosalind Franklin



(b) X-ray diffraction of wet DNA fibers

FIGURE 11.12 **X-ray diffraction of DNA.**

Franklin also studied the fibers' reactions to humid conditions. The diffraction pattern of Franklin's DNA fibers is shown in **Figure 11.12b**. This pattern suggested several structural features of DNA. First, it was consistent with a helical structure. Second, the diameter of the helical structure was too wide to be only a single-stranded helix. Finally, the diffraction pattern indicated that the helix contains about 10 base pairs (bp) per complete turn. These observations were instrumental in solving the structure of DNA.

Chargaff Found That DNA Has a Biochemical Composition in Which the Amount of A Equals T and the Amount of G Equals C

Another piece of information that led to the discovery of the double-helix structure came from the studies of Erwin Chargaff. In the 1940s and 1950s, he pioneered many of the biochemical techniques for the isolation, purification, and measurement of nucleic acids from living cells. This was not a trivial undertaking, because the biochemical composition of living cells is complex. At the time of Chargaff's work, researchers already knew that the building blocks of DNA are nucleotides containing the bases adenine, thymine, guanine, or cytosine. Chargaff analyzed the base composition of DNA, which was isolated from many different species. He expected that the results might provide important clues concerning the structure of DNA.

The data shown in **Table 11.1** are only a small sampling of Chargaff's results. During the late 1940s and early 1950s, Chargaff published many papers concerned with the chemical composition of DNA from biological sources. Hundreds of measurements were made. The compelling observation was that the amount of adenine was similar to that of thymine, and the amount of guanine was similar to cytosine. The idea that the amount of A in DNA equals the amount of T, and the amount of G equals C, is known as **Chargaff's rule.**

These results were not sufficient to propose a model for the structure of DNA. However, they provided the important clue that DNA is structured so that each molecule of adenine interacts with thymine, and each molecule of guanine interacts with cytosine. A DNA structure in which A binds to T, and G to C, would explain the equal amounts of A and T, and G and C observed in Chargaff's experiments. As we will see, this observation became crucial evidence that Watson and Crick used to elucidate the structure of the double helix.

Watson and Crick Deduced the Double-Helical Structure of DNA

Thus far, we have examined key pieces of information used to determine the structure of DNA. In particular, the X-ray diffraction work of Franklin suggested a helical structure composed of two or more strands with 10 bases per turn. In addition, the work of Chargaff indicated that the amount of A equals T, and the amount of G equals C. Furthermore, Watson and Crick were familiar with Pauling's success in using ball-and-stick models to deduce the secondary structure of proteins. With these key observations, they set out to solve the structure of DNA.

TABLE 11.1

Base Content in the DNA from a Variety of Organisms

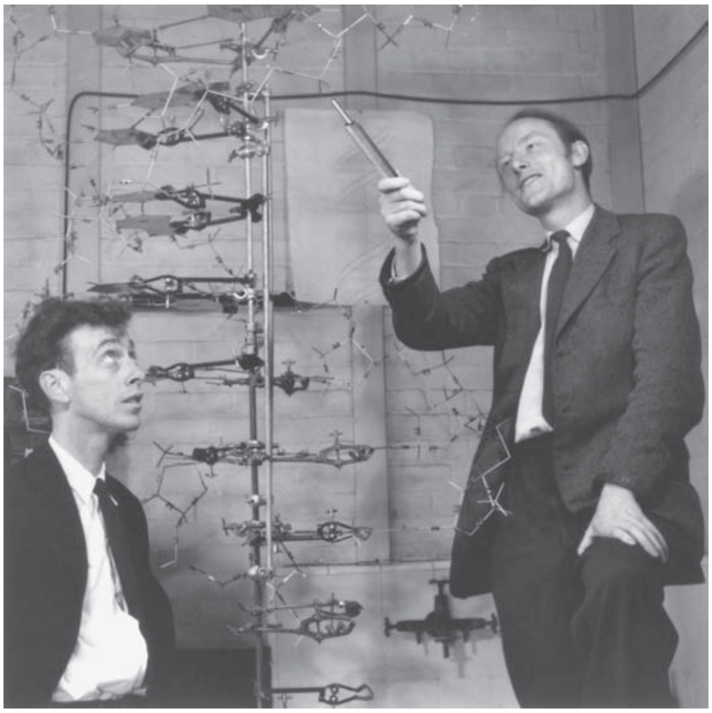
	Percentage of Bases (based on molarity)			
Organism	**Adenine**	**Thymine**	**Guanine**	**Cytosine**
Escherichia coli	26.0	23.9	24.9	25.2
Streptococcus pneumoniae	29.8	31.6	20.5	18.0
Yeast	31.7	32.6	18.3	17.4
Turtle red blood cells	28.7	27.9	22.0	21.3
Salmon sperm	29.7	29.1	20.8	20.4
Chicken red blood cells	28.0	28.4	22.0	21.6
Human liver cells	30.3	30.3	19.5	19.9

*When the base compositions from different tissues within the same species were measured, similar results were obtained. These data were compiled from several sources. See E. Chargaff and J. Davidson, Eds. (1995) *The Nucleic Acids*. Academic Press, New York.

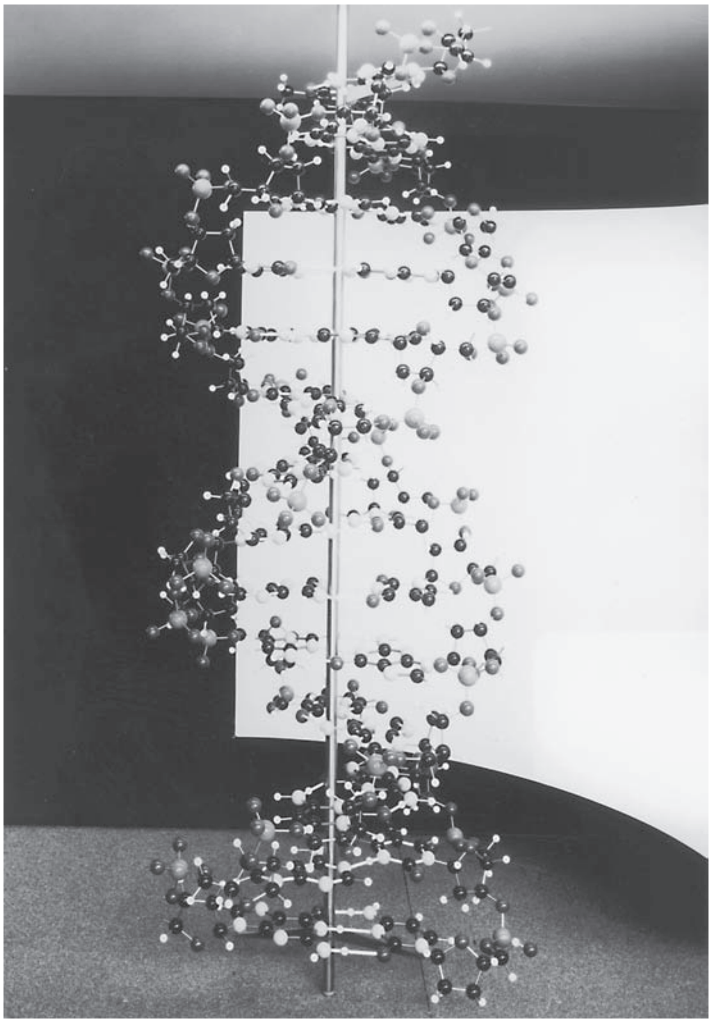
Watson and Crick assumed DNA is composed of nucleotides that are linked together in a linear fashion. They also assumed the chemical linkage between two nucleotides is always the same. With these ideas in mind, they tried to build ball-and-stick models that incorporated the known experimental observations. Because the diffraction pattern suggested the helix must have two (or more) strands, a critical question was, "How could two strands interact?" As discussed in his book *The Double Helix*, James Watson noted that in an early attempt at model building, they considered the possibility that the negatively charged phosphate groups, together with magnesium ions, were promoting an interaction between the backbones of DNA strands. However, more detailed diffraction data were not consistent with this model.

Because the magnesium hypothesis for DNA structure appeared to be incorrect, it was back to the drawing board (or back to the ball-and-stick units) for Watson and Crick. During this time, Rosalind Franklin had produced even clearer X-ray diffraction patterns, which provided greater detail concerning the relative locations of the bases and backbone of DNA. This major breakthrough suggested a two-strand interaction that was helical. In their model building, Watson and Crick's emphasis shifted to models containing the two backbones on the outside of the model, with the bases projecting toward each other. At first, a structure was considered in which the bases form hydrogen bonds with the identical base in the opposite strand (A to A, T to T, G to G, and C to C). However, the model building revealed that the bases could not fit together this way. The final hurdle was overcome when it was realized that the hydrogen bonding of adenine to thymine was structurally similar to that of guanine to cytosine. With an interaction between A and T and between G and C, the ball-and-stick models showed that the two strands would fit together properly. This ball-and-stick model, shown in **Figure 11.13**, was consistent with all of the known data regarding DNA structure.

For their work, Watson, Crick, and Maurice Wilkins were awarded the 1962 Nobel Prize in physiology or medicine. The contribution of Rosalind Franklin to the discovery of the double helix was also critical and has been acknowledged in several books and articles. Franklin was independently trying to solve the structure of DNA. However, Wilkins, who worked in the same laboratory, shared Franklin's X-ray data with Watson and Crick, presumably without her knowledge. This provided important information that helped them solve the structure of DNA, which was published in the journal *Nature* in April 1953. Though she was not given credit in the original publication of the double-helix structure, Franklin's key contribution became known in later years. Unfortunately, however, Rosalind Franklin died in 1958, and the Nobel Prize is not awarded posthumously.



(a) Watson and Crick



(b) Original model of the DNA double helix

FIGURE 11.13 **Watson and Crick and their model of the DNA double helix.** **(a)** James Watson is shown here on the left and Francis Crick on the right. **(b)** The molecular model they originally proposed for the double helix. Each strand contains a sugar–phosphate backbone. In opposite strands, A hydrogen bonds to T, and G hydrogen bonds with C.

11.5 REVIEWING THE KEY CONCEPTS

- Pauling used ball-and-stick models to deduce the secondary structure of a protein (see Figure 11.11).
- Franklin performed X-ray diffraction studies that helped to determine the structure of DNA (see Figure 11.12).
- Chargaff determined that, in DNA, the amount of A equals T and the amount of G equals C (see Table 11.1).
- Watson and Crick deduced the structure of DNA, though they considered incorrect models along the way (see Figure 11.13).

11.5 COMPREHENSION QUESTIONS

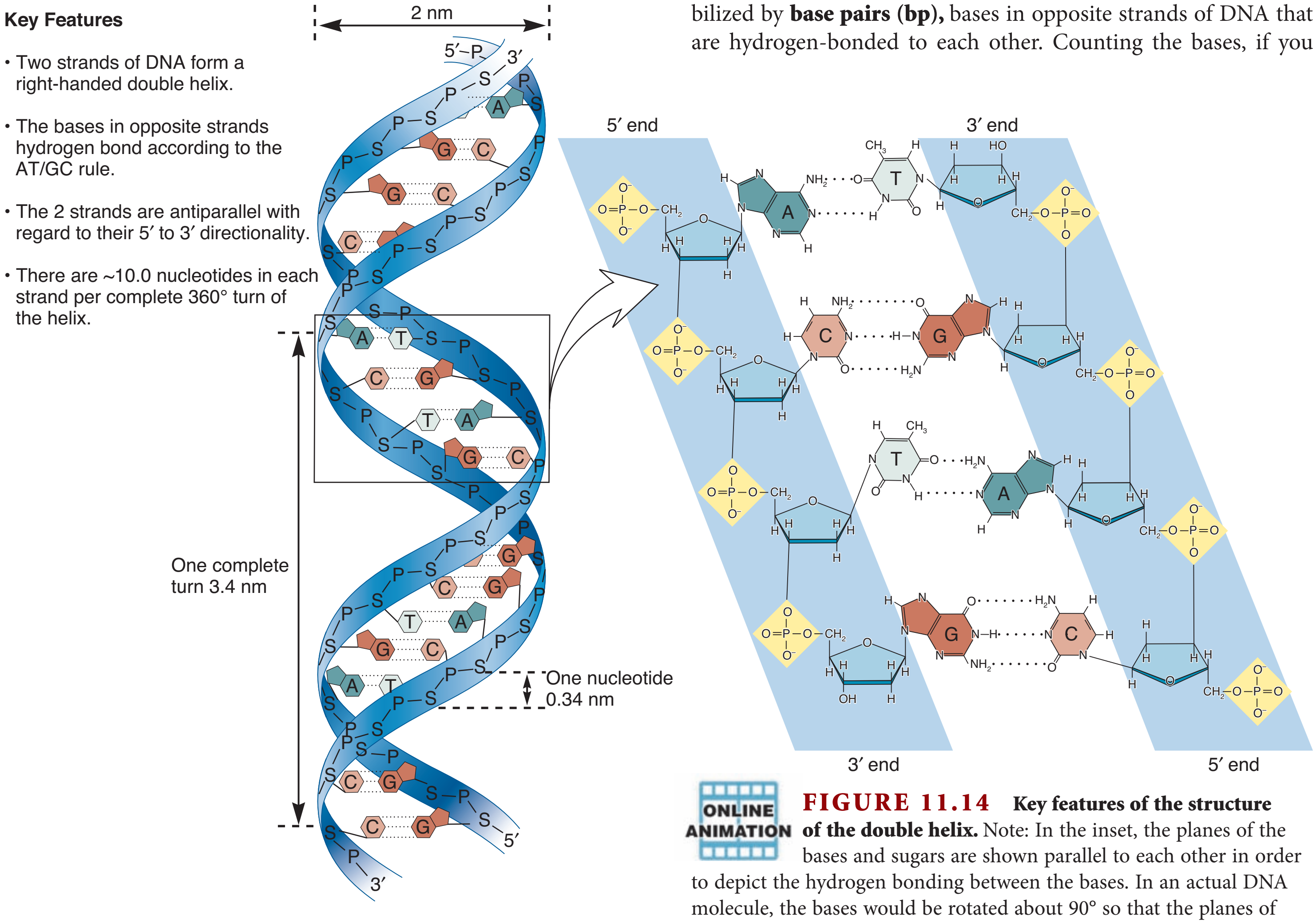
1. Evidence or approaches that led to the discovery of the DNA double helix include
   a. the determination of structures using ball-and-stick models.
   b. the X-ray diffraction data of Franklin.
   c. the base composition data of Chargaff.
   d. all of the above.
2. Chargaff's analysis of the base composition of DNA is consistent with base pairing between
   a. A and G, and T and C.
   b. A and A, G and G, T and T, C and C.
   c. A and T, and G and C.
   d. A and C, and T and G.

11.6 STRUCTURE OF THE DNA DOUBLE HELIX

As we have seen, the discovery of the DNA double helix required different kinds of evidence. In this section, we will examine its structure in greater detail.

The Molecular Structure of the DNA Double Helix Has Several Key Features

The general structural features of the double helix are shown in **Figure 11.14**. In a DNA double helix, two DNA strands are twisted together around a common axis to form a structure that resembles a spiral staircase. This double-stranded structure is stabilized by **base pairs (bp),** bases in opposite strands of DNA that are hydrogen-bonded to each other. Counting the bases, if you


Key Features
• Two strands of DNA form a right-handed double helix.
• The bases in opposite strands hydrogen bond according to the AT/GC rule.
• The 2 strands are antiparallel with regard to their 5′ to 3′ directionality.
• There are ~10.0 nucleotides in each strand per complete 360° turn of the helix.
2 nm
5′
3′
One complete turn 3.4 nm
One nucleotide 0.34 nm
5′ end
3′ end
3′ end
5′ end
A
T
G
C
P
S

ONLINE ANIMATION

FIGURE 11.14 **Key features of the structure of the double helix.** Note: In the inset, the planes of the bases and sugars are shown parallel to each other in order to depict the hydrogen bonding between the bases. In an actual DNA molecule, the bases would be rotated about 90° so that the planes of bases and sugars would be facing each other, as shown in Figure 11.15a.

Concept check: What holds the DNA strands together?

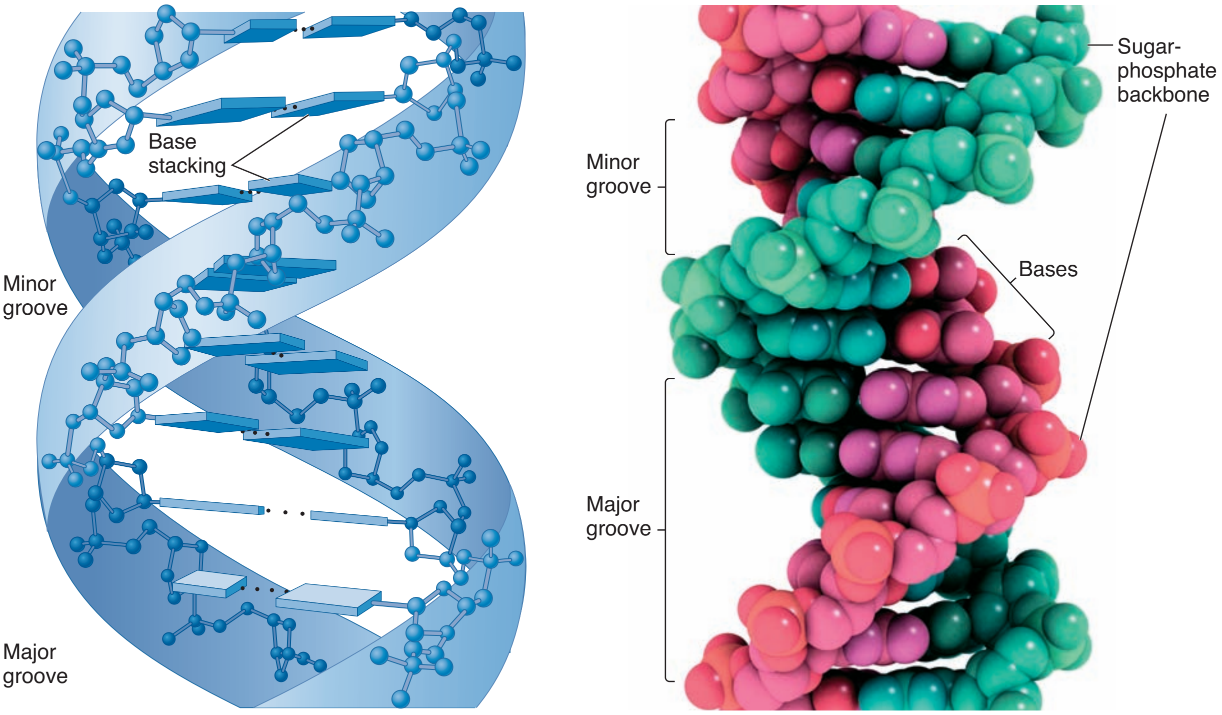
move past 10 bp, you have gone 360° around the backbone. The linear distance of a complete turn is 3.4 nm; each base pair traverses 0.34 nm.

A distinguishing feature of the hydrogen bonding between base pairs is its specificity. An adenine base in one strand hydrogen bonds with a thymine base in the opposite strand, or a guanine base hydrogen bonds with a cytosine. This **AT/GC rule** explains the earlier data of Chargaff showing that the DNA from many organisms contains equal amounts of A and T, and equal amounts of G and C (see Table 11.1). The AT/GC rule indicates that purines (A and G) always bond with pyrimidines (T and C). This keeps the width of the double helix relatively constant. As noted in Figure 11.14, three hydrogen bonds occur between G and C but only two between A and T. For this reason, DNA sequences that have a high proportion of G and C tend to form more stable double-stranded structures.

The AT/GC rule implies that we can predict the sequence in one DNA strand if the sequence in the opposite strand is known. For example, let's consider a DNA strand with the sequence of 5′–ATGGCGGATTT–3′. The opposite strand must be 3′–TACCGCCTAAA–5′. In genetic terms, we would say that these two sequences are **complementary** to each other or that the two sequences exhibit complementarity. In addition, you may have noticed that the sequences are labeled with 5′ and 3′ ends. These numbers designate the direction of the DNA backbones. The direction of DNA strands is depicted in the inset to Figure 11.14. When going from the top of this figure to the bottom, one strand is running in the 5′ to 3′ direction, and the other strand is 3′ to 5′. This opposite orientation of the two DNA strands is referred to as an **antiparallel** arrangement. An antiparallel structure was initially proposed in the model of Watson and Crick.

Figure 11.15a is a schematic model that emphasizes certain molecular features of DNA structure. The bases in this model are depicted as flat rectangular structures that hydrogen bond in pairs. (The hydrogen bonds are the dotted lines.) Although the bases are not actually rectangular, they do form flattened planar structures. Within DNA, the bases are oriented so that the flattened regions are facing each other—an arrangement referred to as base stacking. In other words, if you think of the bases as flat plates, these plates are stacked on top of each other in the double-stranded DNA structure. This is a favorable interaction because the bases are relatively nonpolar and can interact with each other. Along with hydrogen bonding, base stacking is a structural feature that stabilizes the double helix by excluding water molecules, which are polar. The helical structure of the DNA backbone depends on the hydrogen bonding between base pairs and also on base stacking.

By convention, the DNA double helix shown in Figure 11.15a spirals in a direction that is called "right-handed." To understand this terminology, imagine that a double helix is laid on your desk; one end of the helix is close to you, and the other end is at the opposite side of the desk. As it spirals away from you, a right-handed helix turns in a clockwise direction. By comparison, a left-handed helix would spiral in a counterclockwise manner. Both strands in Figure 11.15a spiral in a right-handed direction.



(a) Ball-and-stick model of DNA

(b) Space-filling model of DNA

ONLINE ANIMATION

FIGURE 11.15 **Two models of the double helix.** **(a)** Ball-and-stick model of the double helix. The deoxyribose–phosphate backbone is shown in detail, whereas the bases are depicted as flattened rectangles. **(b)** Space-filling model of the double helix.

Concept check: *Describe the major and minor grooves.*

Figure 11.15b is a space-filling model for DNA in which the atoms are represented by spheres. This model emphasizes the surface features of DNA. Note that the backbone—composed of sugar and phosphate groups—is on the outermost surface. In a living cell, the backbone has the most direct contact with water. In contrast, the bases are more internally located within the double-stranded structure. Biochemists use the term **grooves** to describe the indentations where the atoms of the bases are in contact with the surrounding water. As you travel around the DNA helix, the structure of DNA has two grooves: the **major groove** and the **minor groove.**

As discussed in later chapters, proteins can bind to DNA and affect its conformation and function. For example, some proteins can hydrogen bond to the bases within the major groove. This hydrogen bonding can be very precise so that a protein interacts with a particular sequence of bases. In this way, a protein can recognize a specific gene and affect its ability to be transcribed. We will consider such proteins in Chapters 14, 16, and 17. Alternatively, other proteins bind to the DNA backbone. For example, histone proteins, which are discussed in Chapter 12, form ionic interactions with the negatively charged phosphates in the DNA backbone. The histones are important for the proper compaction of DNA in eukaryotic cells and also play a role in gene transcription.

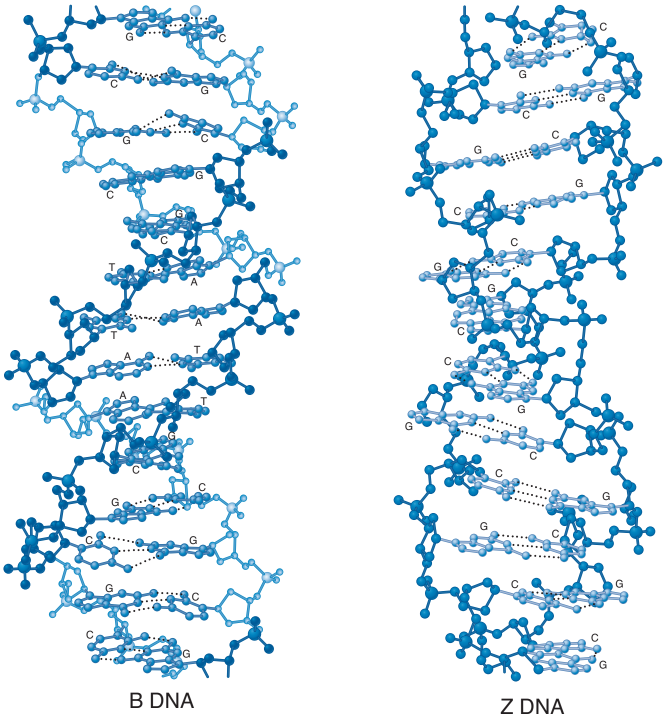
DNA Can Form Alternative Types of Double Helices

The DNA double helix can form different types of structures. **Figure 11.16** compares the structures of **B DNA** and **Z DNA.** The highly detailed structures shown here were deduced by X-ray crystallography on short segments of DNA. B DNA is the predominant form of DNA found in living cells. However, under certain conditions, the two strands of DNA can twist into Z DNA, which differs significantly from B DNA. B DNA is a right-handed helix, whereas Z DNA has a left-handed conformation. In addition, the helical backbone in Z DNA appears to zigzag slightly as it winds itself around the double-helical structure. The numbers of base pairs per 360° turn are 10.0 and 12.0 in B and Z DNA, respectively. In B DNA, the bases tend to be centrally located, and the hydrogen bonds between base pairs occur relatively perpendicular to the central axis. In contrast, the bases in Z DNA are substantially tilted relative to the central axis.

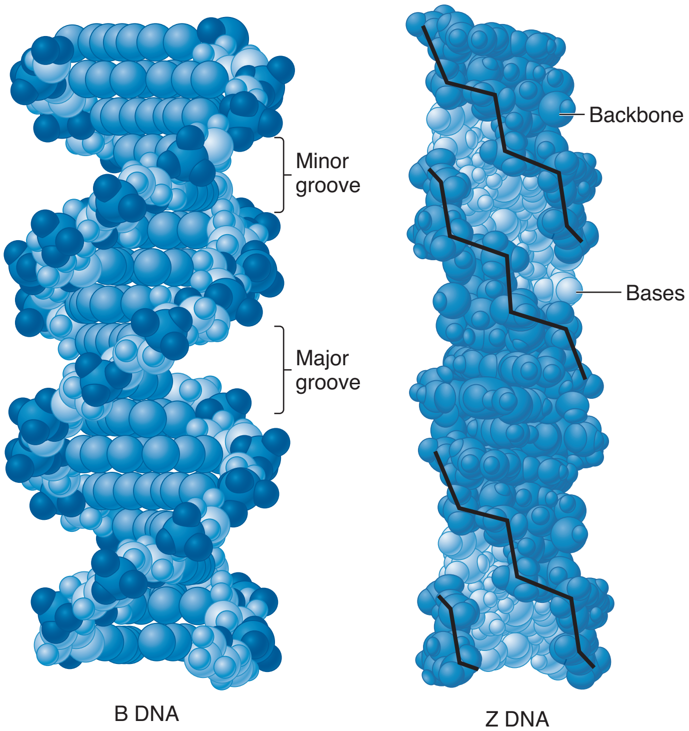
The ability of a double helix to adopt a Z-DNA conformation depends on various factors. At high ionic strength (i.e., high salt concentration), formation of a Z-DNA conformation is favored by a sequence of bases that alternates between purines and pyrimidines. One such sequence is

5′–GCGCGCGCG–3′
3′–CGCGCGCGC–5′

At lower ionic strength, the methylation of cytosine bases can favor Z-DNA formation. Cytosine **methylation** occurs when a cellular enzyme attaches a methyl group ($-CH_3$) to the cytosine base. In addition, negative supercoiling (a topic discussed in Chapter 12) favors the Z-DNA conformation.



(a) Molecular structures



(b) Space-filling models

FIGURE 11.16 Comparison of the structures of B DNA and Z DNA. (a) The highly detailed structures shown here were deduced by X-ray crystallography performed on short segments of DNA. **(b)** Space-filling models of the B-DNA and Z-DNA structures. In the case of Z DNA, the black lines connect the phosphate groups in the DNA backbone. As seen here, they travel along the backbone in a zigzag pattern.

Concept check: What are the structural differences between B DNA and Z DNA?

What is the biological significance of Z DNA? Accumulating evidence suggests a possible biological role for Z DNA in the process of transcription. Recent research has identified cellular proteins that specifically recognize Z DNA. In 2005, Alexander Rich and colleagues reported that the Z-DNA-binding region of one such protein played a role in regulating the transcription of particular genes. In addition, other research has suggested that short regions of Z DNA may play a role in chromosome structure by affecting the level of compaction.

11.6 REVIEWING THE KEY CONCEPTS

- DNA is usually a right-handed double helix in which A hydrogen bonds to T and G hydrogen bonds to C. The two strands are antiparallel and contain about 10 bp per turn. Base stacking also stabilizes the double helix by excluding water (see Figure 11.14).
- The spiral structure of DNA has a major groove and a minor groove (see Figure 11.15).
- B DNA is the major form of DNA found in living cells. Z DNA is an alternative conformation for DNA (see Figure 11.16).

11.6 COMPREHENSION QUESTIONS

1. Which of the following is *not* a feature of the DNA double helix?
   a. It obeys the AT/GC rule.
   b. The DNA strands are antiparallel.
   c. The structure is stabilized by base stacking.
   d. All of the above are correct features of a DNA double helix.
2. A groove in the DNA refers to
   a. the indentations where the atoms of the bases are in contact with the surrounding water.
   b. the interactions between bases in the DNA.
   c. the spiral structure of the DNA.
   d. all of the above.
3. A key difference between B DNA and Z DNA is that
   a. B DNA is right-handed, whereas Z DNA is left-handed.
   b. B DNA obeys the AT/GC rule, whereas Z DNA does not.
   c. Z DNA allows ribose in its structure, whereas B DNA uses deoxyribose.
   d. Z DNA allows uracil in its structure, whereas B DNA uses thymine.

11.7 RNA STRUCTURE

Let's now turn our attention to RNA structure, which has many similarities with DNA structure. The structure of an RNA strand is much like a DNA strand (**Figure 11.17**). Strands of RNA are typically several hundred or several thousand nucleotides in length, which is much shorter than chromosomal DNA. When RNA is made during transcription, the DNA is used as a template to make a copy of single-stranded RNA. In most cases, only one of the two DNA strands is used as a template for RNA synthesis. Therefore, only one complementary strand of RNA is usually made. Nevertheless, relatively short sequences within one RNA molecule or between two separate RNA molecules can form double-stranded regions. RNA double helices are antiparallel and right-handed, with 11 to 12 bp per turn.

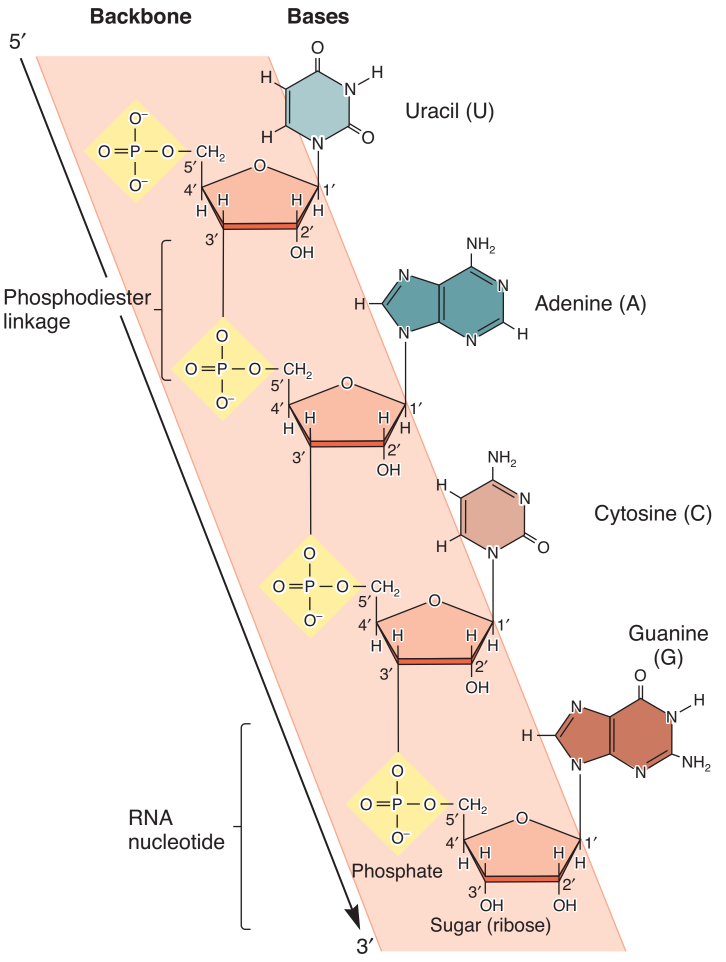


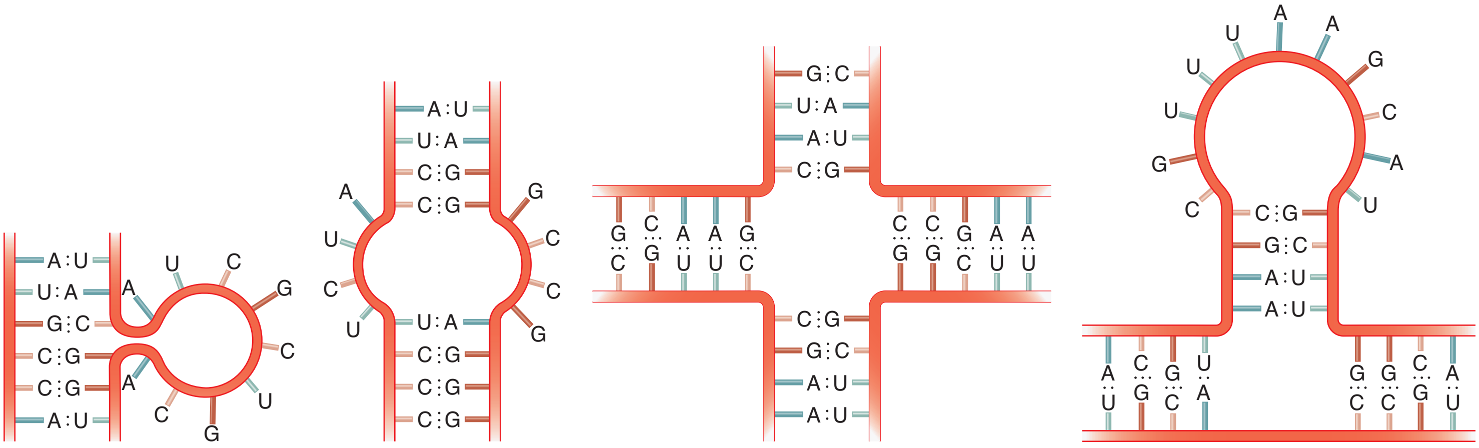
FIGURE 11.17 **A strand of RNA.** This structure is very similar to a DNA strand (see Figure 11.10), except that the sugar is ribose instead of deoxyribose, and uracil is substituted for thymine.

Concept check: What types of bonds hold nucleotides together in an RNA strand?

In places where an RNA molecule is double-stranded, complementary regions form base pairs between A and U and between G and C. Depending on the relative locations of these complementary regions, different types of structural patterns are possible (**Figure 11.18**). These include bulge loops, internal loops, multibranched junctions, and stem-loops (also called hairpins). These structures contain regions of complementarity punctuated by regions of noncomplementarity. As shown in Figure 11.18, the complementary regions are held together by connecting hydrogen bonds, whereas the noncomplementary regions have their bases projecting away from the double-stranded region.

Many factors contribute to the structure of RNA molecules. These include the hydrogen bonding between base pairs, stacking between bases, and hydrogen bonding between bases and backbone regions. In addition, interactions with ions, small molecules, and large proteins may influence RNA structure. In a living cell, the various regions of an RNA molecule fold and interact with each other to produce the three-dimensional structure.

Figure 11.19 depicts the structure of a transfer RNA molecule known as tRNA^{Phe}, which is a tRNA molecule that carries



(a) Bulge loop **(b) Internal loop** **(c) Multibranched junction** **(d) Stem-loop**

FIGURE 11.18 **Possible structures of RNA molecules.** The double-stranded regions are depicted by connecting hydrogen bonds. Loops are noncomplementary regions that are not hydrogen bonded with complementary bases. Double-stranded RNA structures can form within a single RNA molecule or between two separate RNA molecules.

Concept check: *What are the base-pairing rules in RNA?*

the amino acid phenylalanine. It was the first naturally occurring RNA to have its structure elucidated. This RNA molecule has several double-stranded and single-stranded regions.

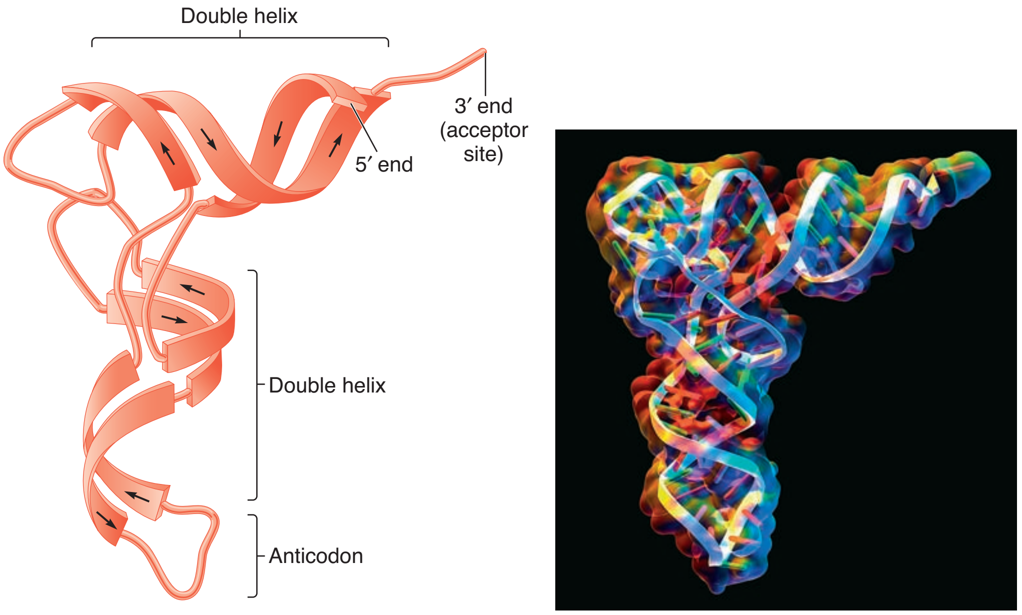
The folding of RNA into a three-dimensional structure is important for its function. For example, as discussed in Chapter 15, a tRNA molecule has two key functional sites—an anticodon and a 3′ acceptor site—that play important roles in translation. In a folded tRNA molecule, these sites are exposed on the surface of the molecule and can perform their roles (see Figure 11.19a). Many other examples are known in which RNA folding is key to its structure and function. These include the folding of ribosomal RNAs (rRNAs), which are important components of ribosomes, and ribozymes, which are RNA molecules with catalytic function.

11.7 REVIEWING THE KEY CONCEPTS

- RNA is also a strand of nucleotides (see Figure 11.17).
- RNA can form double-stranded helical regions and folds into a three-dimensional structure (see Figures 11.18, 11.19).

11.7 COMPREHENSION QUESTIONS

1. A double-stranded region of RNA
    a. forms a helical structure.
    b. obeys the AU/GC rule.
    c. may promote the formation of different structures such as bulge loops and stem-loops.
    d. does all of the above.



(a) Ribbon model **(b) Space-filling model**

FIGURE 11.19 **The structure of tRNA^{Phe}, the transfer RNA molecule that carries phenylalanine.** **(a)** The double-stranded regions of the molecule are shown as antiparallel ribbons. **(b)** A space-filling model of tRNA^{Phe}.

KEY TERMS

Page 225. molecular genetics, DNA (deoxyribonucleic acid), RNA (ribonucleic acid)
Page 226. transformation
Page 228. DNase, RNase, protease, bacteriophage, phage
Page 229. lysis
Page 231. nucleic acids, nucleotides, strand, double helix
Page 232. deoxyribose, ribose, purines, pyrimidines, adenine (A), guanine (G), thymine (T), cytosine (C), uracil (U), nucleoside
Page 233. polynucleotides
Page 234. phosphodiester linkage, backbone, directionality
Page 235. Chargaff's rule
Page 237. base pairs (bp)
Page 238. AT/GC rule, complementary, antiparallel
Page 239. grooves, major groove, minor groove, B DNA, Z DNA, methylation

CHAPTER SUMMARY

- Molecular genetics is the study of DNA structure and function at the molecular level.

11.1 Identification of DNA as the Genetic Material

- To fulfill its role, a genetic material must meet four criteria: information, transmission, replication, and variation.
- Griffith showed that the genetic material from type S bacteria could transform type R bacteria into type S (see Figure 11.1).
- Avery, MacLeod, and McCarty discovered that the transforming substance is DNA (see Figure 11.2).
- Hershey and Chase determined that the genetic material of T2 phage is DNA (see Figures 11.3–11.5).
- Viruses may use DNA or RNA as their genetic material.

11.2 Overview of DNA and RNA Structure

- DNA and RNA are types of nucleic acids.
- In DNA, nucleotides are linked together to form strands, which then form a double helix that is found within chromosomes (see Figure 11.6).

11.3 Nucleotide Structure

- A nucleotide is composed of one or more phosphates, a sugar, and a base. The purine bases are adenine and guanine, whereas the pyrimidine bases are thymine (DNA only), cytosine, and uracil (RNA only). The sugar deoxyribose is found in DNA, whereas ribose is in RNA (see Figures 11.7–11.8).
- A nucleoside contains a sugar and a base, whereas a nucleotide contains a sugar, a base, and one or more phosphate groups (see Figure 11.9).

11.4 Structure of a DNA Strand

- In a DNA strand, nucleotides are covalently attached to one another via phosphodiester linkages (see Figure 11.10).

11.5 Discovery of the Double Helix

- Pauling used ball-and-stick models to deduce the secondary structure of a protein (see Figure 11.11).
- Franklin performed X-ray diffraction studies that helped to determine the structure of DNA (see Figure 11.12).
- Chargaff determined that, in DNA, the amount of A equals T and the amount of G equals C (see Table 11.1).
- Watson and Crick deduced the structure of DNA, though they considered incorrect models along the way (see Figure 11.13).

11.6 Structure of the DNA Double Helix

- DNA is usually a right-handed double helix in which A hydrogen bonds to T and G hydrogen bonds to C. The two strands are antiparallel and contain about 10 bp per turn. Base stacking also stabilizes the double helix by excluding water (see Figure 11.14).
- The spiral structure of DNA has a major groove and a minor groove (see Figure 11.15).
- B DNA is the major form of DNA found in living cells. Z DNA is an alternative conformation for DNA (see Figure 11.16).

11.7 RNA Structure

- RNA is also a strand of nucleotides (see Figure 11.17).
- RNA can form double-stranded helical regions, and it folds into a three-dimensional structure (see Figures 11.18, 11.19).

PROBLEM SETS & INSIGHTS

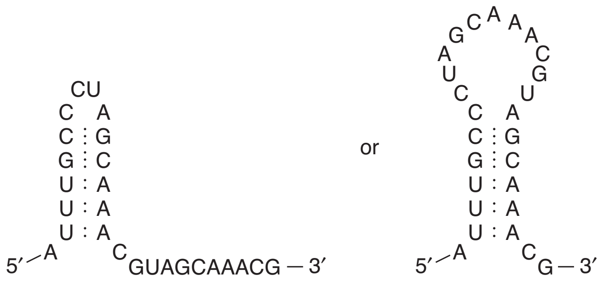
Solved Problems

S1. A hypothetical sequence at the beginning of an mRNA molecule is

5′–AUUUGCCCUAGCAAACGUAGCAAACG...rest of the coding sequence

Using two out of the three underlined sequences, draw two possible models for potential stem-loop structures at the 5′ end of this mRNA.

Answer:



S2. Describe the previous experimental evidence that led Watson and Crick to the discovery of the DNA double helix.

Answer:

1. The chemical structure of single nucleotides was understood by the 1950s.
2. Watson and Crick assumed DNA is composed of nucleotides linked together in a linear fashion to form a strand. They also assumed the chemical linkage between two nucleotides is always the same.
3. Franklin's diffraction patterns suggested several structural features. First, it was consistent with a helical structure. Second, the diameter of the helical structure was too wide to be only a single-stranded helix. Finally, the line spacing on the diffraction pattern indicated the helix contains about 10 bp per complete turn.
4. In the chemical analysis of the DNA from different species, the work of Chargaff indicated that the amount of adenine equaled the amount of thymine, and the amount of cytosine equaled that of guanine.
5. In the early 1950s, Linus Pauling proposed that regions of proteins can fold into a secondary structure. To discover this, Pauling built large models by linking together simple ball-and-stick units. In this way, he could determine if atoms fit together properly in a complicated three-dimensional structure. A similar approach was used by Watson and Crick to solve the structure of the DNA double helix.

S3. Within living cells, many different proteins play important functional roles by binding to DNA and RNA. As described throughout your textbook, the dynamic interactions between nucleic acids and proteins lie at the heart of molecular genetics. Some proteins bind to DNA (or RNA) but not in a sequence-specific manner. For example, histones are proteins important in the formation of chromosome structure. In this case, the positively charged histone proteins bind to the negatively charged phosphate groups in DNA. In addition, several other proteins interact with DNA but do not require a specific nucleotide sequence to carry out their function. For example, DNA polymerase, which catalyzes the synthesis of new DNA strands, does not bind to DNA in a sequence-dependent manner. By comparison, many other proteins do interact with nucleic acids in a sequence-dependent fashion. This means that a specific sequence of bases can provide a structure that is recognized by a particular protein. Throughout the textbook, the functions of many of these proteins will be described. Some examples include transcription factors that affect the rate of transcription, proteins that bind to centromeres, and proteins that bind to origins of replication. With regard to the three-dimensional structure of DNA, where would you expect DNA-binding proteins to bind if they recognize a specific base sequence? What about DNA-binding proteins that do not recognize a base sequence?

Answer: DNA-binding proteins that recognize a base sequence must bind into a major or minor groove of the DNA, which is where the bases are accessible to a DNA-binding protein. Most DNA-binding proteins, which recognize a base sequence, fit into the major groove. By comparison, other DNA-binding proteins, such as histones, which do not recognize a base sequence, bind to the DNA backbone.

S4. The formation of a double-stranded structure must obey the rule that adenine hydrogen bonds to thymine (or uracil) and cytosine hydrogen bonds to guanine. Based on your previous understanding of genetics (from this course or a general biology course), discuss reasons why complementarity is an important feature of DNA and RNA structure and function.

Answer: Note: Many of the topics described here are discussed in Chapters 12 through 15. One way that complementarity underlies function is that it provides the basis for the synthesis of new strands of DNA and RNA. During replication, the synthesis of the new DNA strands occurs in such a way that adenine hydrogen bonds to thymine, and cytosine hydrogen bonds to guanine. In other words, the molecular feature of a complementary double-stranded structure makes it possible to produce exact copies of DNA. Likewise, the ability to transcribe DNA into RNA is based on complementarity. During transcription, one strand of DNA is used as a template to make a complementary strand of RNA.

In addition to the synthesis of new strands of DNA and RNA, complementarity is important in other ways. As mentioned in this chapter, the folding of RNA into a particular structure is driven by the hydrogen bonding of complementary regions. This event is necessary to produce functionally active tRNA molecules. Likewise, stem-loop structures also occur in other types of RNA. For example, the rapid formation of stem-loop structures is known to occur as RNA is being transcribed and to affect the termination of transcription.

A third way that complementarity can be functionally important is that it can promote the interaction of two separate RNA molecules. During translation, codons in mRNA bind to the anticodons in tRNA (see Chapter 15). This binding is due to complementarity. For example, if a codon is 5′–AGG–3′, the anticodon is 3′–UCC–5′. This type of specific interaction between codons and anticodons is an important step that enables the nucleotide sequence in mRNA to code for an amino acid sequence within a protein. In addition, many other examples of RNA–RNA interactions are known and will be described throughout this textbook.

Conceptual Questions

C1. What is the meaning of the term *genetic material*?

C2. After the DNA from type S bacteria is exposed to type R bacteria, list all of the steps that you think must occur for the bacteria to start making a capsule.

C3. Look up the meaning of the word *transformation* in a dictionary and explain whether it is an appropriate word to describe the transfer of genetic material from one organism to another.

C4. What are the building blocks of a nucleotide? With regard to the 5′ and 3′ positions on a sugar molecule, how are nucleotides linked together to form a strand of DNA?

C5. Draw the structure of guanine, guanosine, and deoxyguanosine triphosphate.

C6. Draw the structure of a phosphodiester linkage.

C7. Describe how bases interact with each other in the double helix. This discussion should address the issues of complementarity, hydrogen bonding, and base stacking.

C8. If one DNA strand is 5′–GGCATTACACTAGGCCT–3′, what is the sequence of the complementary strand?

C9. What is meant by the term *DNA sequence*?

C10. Make a side-by-side drawing of two DNA helices, one with 10 bp per 360° turn and the other with 15 bp per 360° turn.

C11. Discuss the differences in the structural features of B DNA and Z DNA.

C12. What parts of a nucleotide (namely, phosphate, sugar, and/or bases) occupy the major and minor grooves of double-stranded DNA, and what parts are found in the DNA backbone? If a DNA-binding protein does not recognize a specific nucleotide sequence, do you expect that it recognizes the major groove, the minor groove, or the DNA backbone? Explain.

C13. List the structural differences between DNA and RNA.

C14. Draw the structure of deoxyribose and number the carbon atoms. Describe the numbering of the carbon atoms in deoxyribose with regard to the directionality of a DNA strand. In a DNA double helix, what does the term *antiparallel* mean?

C15. Write out a sequence of an RNA molecule that could form a stem-loop with 24 nucleotides in the stem and 16 nucleotides in the loop.

C16. Compare the structural features of a double-stranded RNA structure with those of a DNA double helix.

C17. Which of the following DNA double helices would be more difficult to separate into single-stranded molecules by treatment with heat, which breaks hydrogen bonds?

A. GGCGTACCAGCGCAT
CCGCATGGTCGCGTA

B. ATACGATTTACGAGA
TATGCTAAATGCTCT

Explain your choice.

C18. What structural feature allows DNA to store information?

C19. Discuss the structural significance of complementarity in DNA and in RNA.

C20. An organism has a G + C content of 64% in its DNA. What are the percentages of A, T, G, and C?

C21. A DNA-binding protein recognizes the following double-stranded sequence:

5′–GCCCGGGC–3′
3′–CGGGCCCG–5′

This type of double-stranded structure could also occur within the stem region of an RNA stem-loop molecule. Discuss the structural differences between RNA and DNA that might prevent this DNA-binding protein from recognizing a double-stranded RNA molecule.

C22. Within a protein, certain amino acids are positively charged (e.g., lysine and arginine), some are negatively charged (e.g., glutamate and aspartate), some are polar but uncharged, and some are nonpolar. If you knew that a DNA-binding protein was recognizing the DNA backbone rather than base sequences, which amino acids in the protein would be good candidates for interacting with the DNA?

C23. A double-stranded DNA molecule contains 560 nucleotides. How many complete turns would be found in this double helix?

C24. As the minor and major grooves of the DNA wind around a DNA double helix, do they ever intersect each other, or do they always run parallel to each other?

C25. What chemical group (phosphate group, hydroxyl group, or a nitrogenous base) is found at the 3′ end of a DNA strand? What group is found at the 5′ end?

C26. The base composition of an RNA virus was analyzed and found to be 14.1% A, 14.0% U, 36.2% G, and 35.7% C. Would you conclude that the viral genetic material is single-stranded RNA or double-stranded RNA?

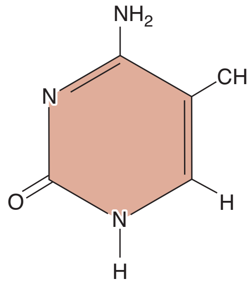
C27. The genetic material found within some viruses is single-stranded DNA. Would this genetic material contain equal amounts of A and T and equal amounts of G and C?

C28. A medium-sized human chromosome contains about 100 million base pairs. If the DNA were stretched out in a linear manner, how long would it be?

C29. A double-stranded DNA molecule is 1 cm long, and the percentage of adenine is 15%. How many cytosines would be found in this DNA molecule?

C30. Could single-stranded DNA form a stem-loop structure? Why or why not?

C31. As described in Chapter 17, the methylation of cytosine bases can have an important effect on gene expression. For example, the methylation of cytosines may inhibit the transcription of genes. A methylated cytosine base has the following structure:



Would you expect the methylation of cytosine to affect the hydrogen bonding between cytosine and guanine in a DNA double helix? Why or why not? (Hint: See Figure 11.14 for help.) Take a look at Solved Problem S3 and speculate about how methylation could affect gene expression.

C32. An RNA molecule has the following sequence:

Region 1 Region 2 Region 3
5′–CAUCCAUCCAUUCCCCAUCCGAUAAGGGGAAUGGAUCCGAAUGGAUAAC–3′

Parts of region 1 can form a stem-loop with region 2 and with region 3. Can region 1 form a stem-loop with region 2 and region 3 at the same time? Why or why not? Which stem-loop would you predict to be more stable: a region 1/region 2 interaction or a region 1/region 3 interaction? Explain your choice.

Application and Experimental Questions

E1. Genetic material acts as a blueprint for an organism’s traits. Explain how the experiments of Griffith indicated that genetic material was being transferred to the type R bacteria.

E2. With regard to the experiment described in Figure 11.2, answer the following:

A. List several possible reasons why only a small percentage of the type R bacteria was converted to type S.

B. Explain why an antibody must be used to remove the bacteria that are not transformed. What would the results look like, in all five cases, if the antibody/centrifugation step had not been included in the experimental procedure?

C. The DNA extract was treated with DNase, RNase, or protease. Why was this done? (In other words, what were the researchers trying to demonstrate?)

E3. An interesting trait that some bacteria exhibit is resistance to killing by antibiotics. For example, certain strains of bacteria are resistant to tetracycline, whereas other strains are sensitive to tetracycline. Describe an experiment to demonstrate that tetracycline resistance is an inherited trait encoded by the DNA of the resistant strain.

E4. With regard to the experiment of Figure 11.5, answer the following:

A. Provide possible explanations why some of the DNA is in the supernatant.

B. Plot the results if the radioactivity in the pellet, rather than in the supernatant, had been measured.

C. Why were ^{32}P and ^{35}S chosen as radioisotopes to label the phages?

D. List possible reasons why less than 100% of the phage protein was removed from the bacterial cells during the shearing process.

E5. Does the experiment of Figure 11.5 rule out the possibility that RNA is the genetic material of T2 phage? Explain your answer. If it does not, could you modify the approach of Hershey and Chase to show that DNA and not RNA is the genetic material of T2 bacteriophage? *Note:* It is possible to specifically label DNA or RNA by providing bacteria with radiolabeled thymine or uracil, respectively.

E6. In this chapter, we considered two experiments—one by Avery, MacLeod, and McCarty and the second by Hershey and Chase–that indicated DNA is the genetic material. Discuss the strengths and weaknesses of the two approaches. Which experimental approach did you find the most convincing? Why?

E7. The type of model building used by Pauling and Watson and Crick involved the use of ball-and-stick units. Now we can do model building on a computer screen. Even though you may not be familiar with this approach, discuss potential advantages of using computers in molecular model building.

E8. With regard to Chargaff’s data described in Table 11.1, would his experiments have been convincing if they had been done on only one species? Discuss.

E9. Gierer and Schramm exposed plant tissue to purified RNA from tobacco mosaic virus, and the plants developed the same types of lesions as if they were exposed to the virus itself. What would be the results if the RNA was treated with DNase, RNase, or protease prior to its exposure to the plant tissue?

Questions for Student Discussion/Collaboration

1. Try to propose structures for a genetic material that are substantially different from the double helix. Remember that the genetic material must have a way to store information and a way to be faithfully replicated.
2. How might you provide evidence that DNA is the genetic material in mice?

Answers to Comprehension Questions

11.1: d, b
11.2: c
11.3: c, d
11.4: a
11.5: d, c
11.6: d, a, a
11.7: d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

12

CHAPTER OUTLINE

12.1 Organization of Sites Along Bacterial Chromosomes

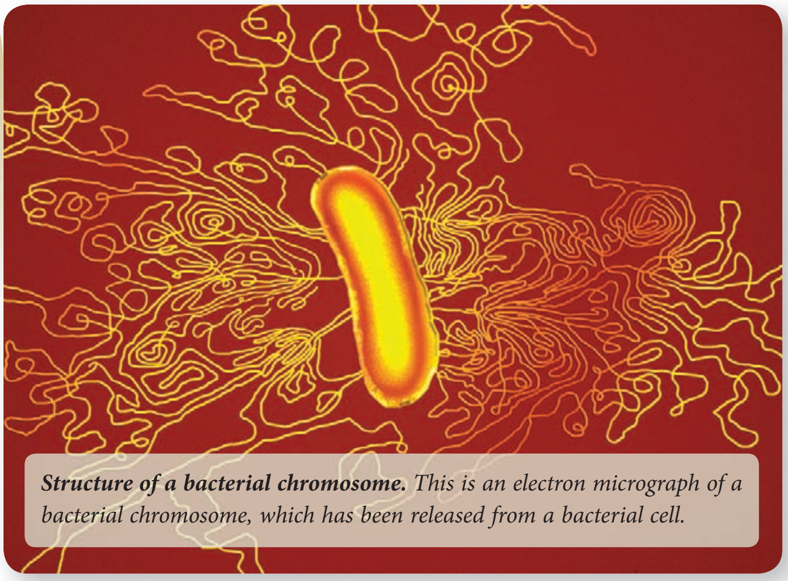
12.2 Structure of Bacterial Chromosomes

12.3 Organization of Sites Along Eukaryotic Chromosomes

12.4 Sizes of Eukaryotic Genomes and Repetitive Sequences

12.5 Structure of Eukaryotic Chromosomes in Nondividing Cells

12.6 Structure of Eukaryotic Chromosomes During Cell Division



Structure of a bacterial chromosome. *This is an electron micrograph of a bacterial chromosome, which has been released from a bacterial cell.*

MOLECULAR STRUCTURE AND ORGANIZATION OF CHROMOSOMES

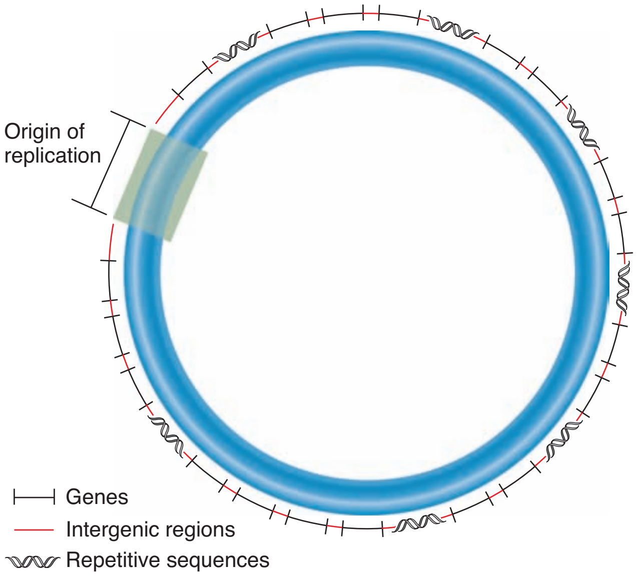
Chromosomes are the structures within cells that contain the genetic material. The term **genome** describes all the types of genetic material that an organism has. For bacteria, the genome is typically a single circular chromosome. For eukaryotes, the nuclear genome refers to one complete set of chromosomes that resides in the cell nucleus. In other words, the haploid complement of chromosomes is considered a nuclear genome. Eukaryotes have a mitochondrial genome, and plants also have a chloroplast genome. Unless otherwise noted, the term *eukaryotic genome* refers to the nuclear genome.

The primary function of the genetic material is to store the information needed to produce the characteristics of an organism. As we saw in Chapter 11, the sequence of bases in a DNA molecule can store information. To fulfill their role at the molecular level, chromosomal sequences facilitate four important processes: (1) the synthesis of RNA and cellular proteins, (2) the replication of chromosomes, (3) the proper segregation of chromosomes, and (4) the compaction of chromosomes so they can fit within living cells. In this chapter, we will examine two features of bacterial and eukaryotic chromosomes. First, we will consider the general organization of functional sites along a chromosome. Second, we will consider the structures of chromosomes. As you will learn, different mechanisms are responsible for making chromosomes more compact so they can fit within living cells.

12.1 ORGANIZATION OF SITES ALONG BACTERIAL CHROMOSOMES

Bacterial chromosomal DNA is usually a circular molecule, though some bacteria have linear chromosomes. A typical chromosome is a few million base pairs (bp) in length. For example, the chromosome of *Escherichia coli* has approximately 4.6 million bp, and the *Haemophilus influenzae* chromosome has roughly 1.8 million bp. A bacterial chromosome commonly has a few thousand different genes. These genes are interspersed throughout the entire chromosome (**Figure 12.1**). **Structural genes**—nucleotide sequences that encode proteins—account for the majority of bacterial DNA. The nontranscribed regions of DNA located between adjacent genes are termed **intergenic regions.**

Other sequences in chromosomal DNA influence DNA replication, gene transcription, and chromosome structure. For



Key features:

- Most, but not all, bacterial species contain circular chromosomal DNA.
- A typical chromosome is a few million base pairs in length.
- Most bacterial species contain a single type of chromosome, but it may be present in multiple copies.
- A few thousand different genes are interspersed throughout the chromosome. The short regions between adjacent genes are called intergenic regions.
- One origin of replication is required to initiate DNA replication.
- Repetitive sequences may be interspersed throughout the chromosome.

FIGURE 12.1 **Organization of sequences in bacterial chromosomal DNA.**

Concept check: *What types of sequences constitute most of a bacterial genome?*

example, bacterial chromosomes have one **origin of replication,** a sequence that is a few hundred nucleotides in length. This nucleotide sequence functions as an initiation site for the assembly of several proteins required for DNA replication. Also, a variety of **repetitive sequences** have been identified in many bacterial species. These sequences are found in multiple copies and are usually interspersed within the intergenic regions throughout the bacterial chromosome. Repetitive sequences may play a role in a variety of genetic processes, including DNA folding, DNA replication, gene regulation, and genetic recombination. As discussed in Chapter 21, some repetitive sequences are transposable elements that can move throughout the genome. Figure 12.1 summarizes the key features of sequence organization within bacterial chromosomes.

12.1 REVIEWING THE KEY CONCEPTS

- Bacterial chromosomes are typically circular and carry an origin of replication and a few thousand genes (see Figure 12.1).

12.1 COMPREHENSION QUESTIONS

1. A bacterial chromosome typically contains
   a. a few thousand genes.
   b. one origin of replication.
   c. some repetitive sequences.
   d. all of the above.

12.2 STRUCTURE OF BACTERIAL CHROMOSOMES

Inside a bacterial cell, a chromosome is highly compacted and found within a region of the cell known as a **nucleoid.** Although bacteria usually contain a single type of chromosome, more than one copy of that chromosome may be found within one bacterial cell. Depending on the growth conditions, bacteria may have one to four identical chromosomes per cell. In addition, the number of copies varies depending on the bacterial species. As shown in **Figure 12.2**, each chromosome occupies its own distinct nucleoid region within the cell. Unlike the eukaryotic nucleus, the bacterial nucleoid is not a separate cellular compartment surrounded by a membrane. Rather, the DNA in a nucleoid is in direct contact with the cytoplasm of the cell. In this section, we will explore the structure of bacterial chromosomes and how they are compacted to fit within the nucleoid region.

The Formation of Chromosomal Loops Helps Make the Bacterial Chromosome More Compact

To fit within the nucleoid of a bacterial cell, the chromosomal DNA must be compacted about 1000-fold. Part of this

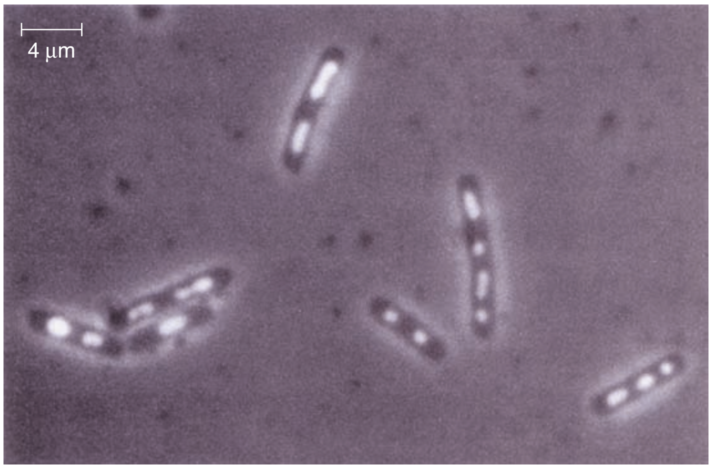
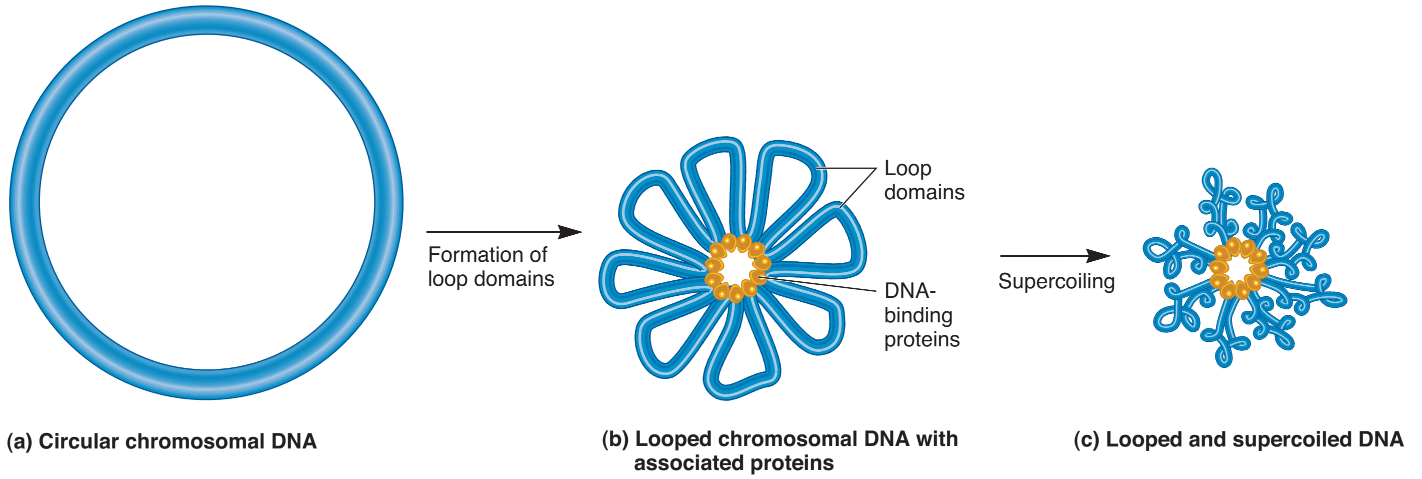


FIGURE 12.2 **The localization of nucleoids within *Bacillus subtilis* bacteria.** The nucleoids are fluorescently labeled and seen as bright, oval-shaped regions within the bacterial cytoplasm. Note that two or more nucleoids are found within each cell. Some of the cells seen here are in the process of dividing.

Concept check: *How many nucleoids are in the bacterial cell on the far right?*


Formation of loop domains
Loop domains
DNA-binding proteins
Supercoiling
(a) Circular chromosomal DNA
(b) Looped chromosomal DNA with associated proteins
(c) Looped and supercoiled DNA

ONLINE ANIMATION

FIGURE 12.3 **The formation of loop domains and DNA supercoiling within the bacterial chromosome.** To promote compaction, **(a)** the large, circular chromosomal DNA is organized into **(b)** smaller loop domains with associated proteins. The looped chromosomal DNA becomes much more compacted due to **(c)** supercoiling within the loops.

Concept check: *What holds the loop domains in place?*

compaction process involves the formation of **loop domains** within the bacterial chromosome (**Figure 12.3a, b**). As its name suggests, a loop domain is a segment of chromosomal DNA folded into a structure that resembles a loop. DNA-binding proteins anchor the base of the loops in place. The number of loops varies according to the size of the bacterial chromosome and the species. In *E. coli*, a chromosome has 50 to 100 loop domains with about 40,000 to 80,000 bp of DNA in each loop. This looped structure compacts the circular chromosome about 10-fold.

DNA Supercoiling Further Compacts the Bacterial Chromosome

Because DNA is a long, thin molecule, twisting forces can dramatically change its conformation. This effect is similar to twisting a rubber band. If twisted in one direction, a rubber band eventually coils itself into a compact structure as it absorbs the energy applied by the twisting motion. Because the two strands within DNA already coil around each other, the formation of additional coils due to twisting forces is referred to as **DNA supercoiling** (**Figure 12.3c**).

How do twisting forces affect DNA structure? **Figure 12.4** illustrates four possibilities. In Figure 12.4a, a double-stranded DNA molecule with five complete turns is anchored between two plates. In this hypothetical example, the ends of the DNA molecule cannot rotate freely. Both underwinding and overwinding of the DNA double helix can induce supercoiling of the helix. Because B DNA is a right-handed helix, underwinding is a left-handed twisting motion, and overwinding is a right-handed twist. Along the left side of Figure 12.4, one of the plates has been given a turn in the direction that tends to unwind the helix. As the helix absorbs this force, two things can happen. The underwinding motion can cause fewer turns (Figure 12.4b) or cause a negative supercoil to form (Figure 12.4c). On the right side of Figure 12.4, one of the plates has been given a right-handed turn, which overwinds the double helix. This can lead to either more turns (Figure 12.4d) or the formation of a positive supercoil (Figure 12.4e). The DNA conformations shown in Figure 12.4a, c, and e differ only with regard to supercoiling. These three DNA conformations are referred to as **topoisomers** of each other. The DNA conformations shown in Figure 12.4b and d are not structurally favorable and do not occur in living cells.

Chromosome Function Is Influenced by DNA Supercoiling

The chromosomal DNA in bacteria is negatively supercoiled. In the chromosome of *E. coli*, about one negative supercoil occurs per 40 turns of the double helix. Negative supercoiling has several important consequences. As already mentioned, the supercoiling of chromosomal DNA makes it much more compact (see Figure 12.3c). Therefore, supercoiling helps to greatly decrease the size of the bacterial chromosome. In addition, negative supercoiling also affects DNA function. To understand how, remember that negative supercoiling is due to an underwinding force on the DNA. Therefore, negative supercoiling creates tension on the DNA strands that may be released by DNA strand separation (**Figure 12.5**). Although most of the chromosomal DNA is negatively supercoiled and compact, the force of negative supercoiling may promote DNA strand separation in small regions. This enhances genetic activities such as replication and transcription, which require separation of the DNA strands.

How does bacterial DNA become supercoiled? In 1976, Martin Gellert and colleagues discovered the enzyme **DNA gyrase,** also known as topoisomerase II. This enzyme, which contains four subunits (two A and two B subunits), introduces negative supercoils (or relaxes positive supercoils) using energy from ATP (**Figure 12.6a**). To alter supercoiling, DNA gyrase has two sets of jaws (also called gates) that allow it to grab onto two regions of DNA. One of the DNA regions is grabbed by the lower

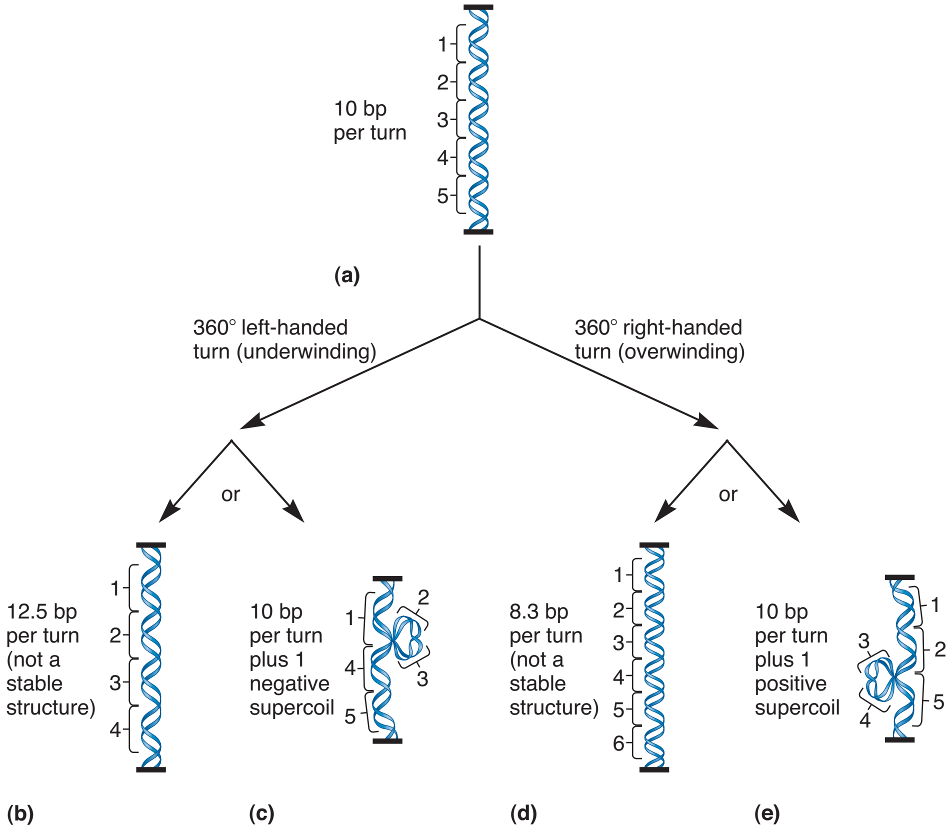

10 bp
per turn
1
2
3
4
5
(a)
360° left-handed
turn (underwinding)
360° right-handed
turn (overwinding)
or
or
12.5 bp
per turn
(not a
stable
structure)
1
2
3
4
(b)
10 bp
per turn
plus 1
negative
supercoil
1
2
3
4
5
(c)
8.3 bp
per turn
(not a
stable
structure)
1
2
3
4
5
6
(d)
10 bp
per turn
plus 1
positive
supercoil
1
2
3
4
5
(e)

FIGURE 12.4 **Schematic representation of DNA supercoiling.** In this example, the DNA in **(a)** is anchored between two plates and given a twist as noted by the arrows. A left-handed twist (underwinding) could produce either **(b)** fewer turns or **(c)** a negative supercoil. A right-handed twist (overwinding) produces **(d)** more turns or **(e)** a positive supercoil. The structures shown in (b) and (d) are unstable.

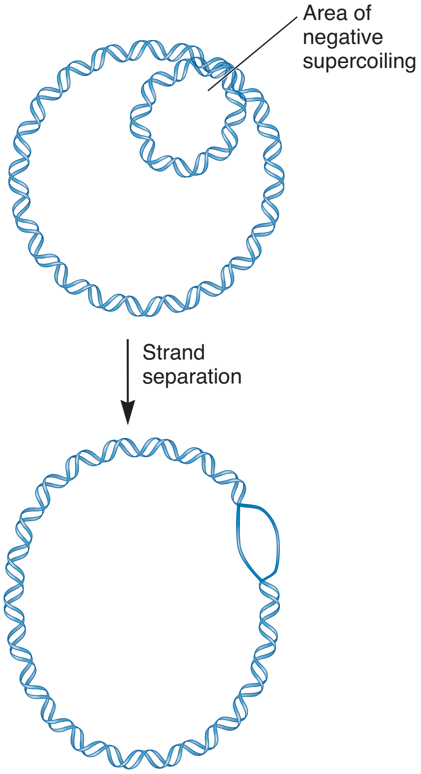

Area of
negative
supercoiling
Strand
separation

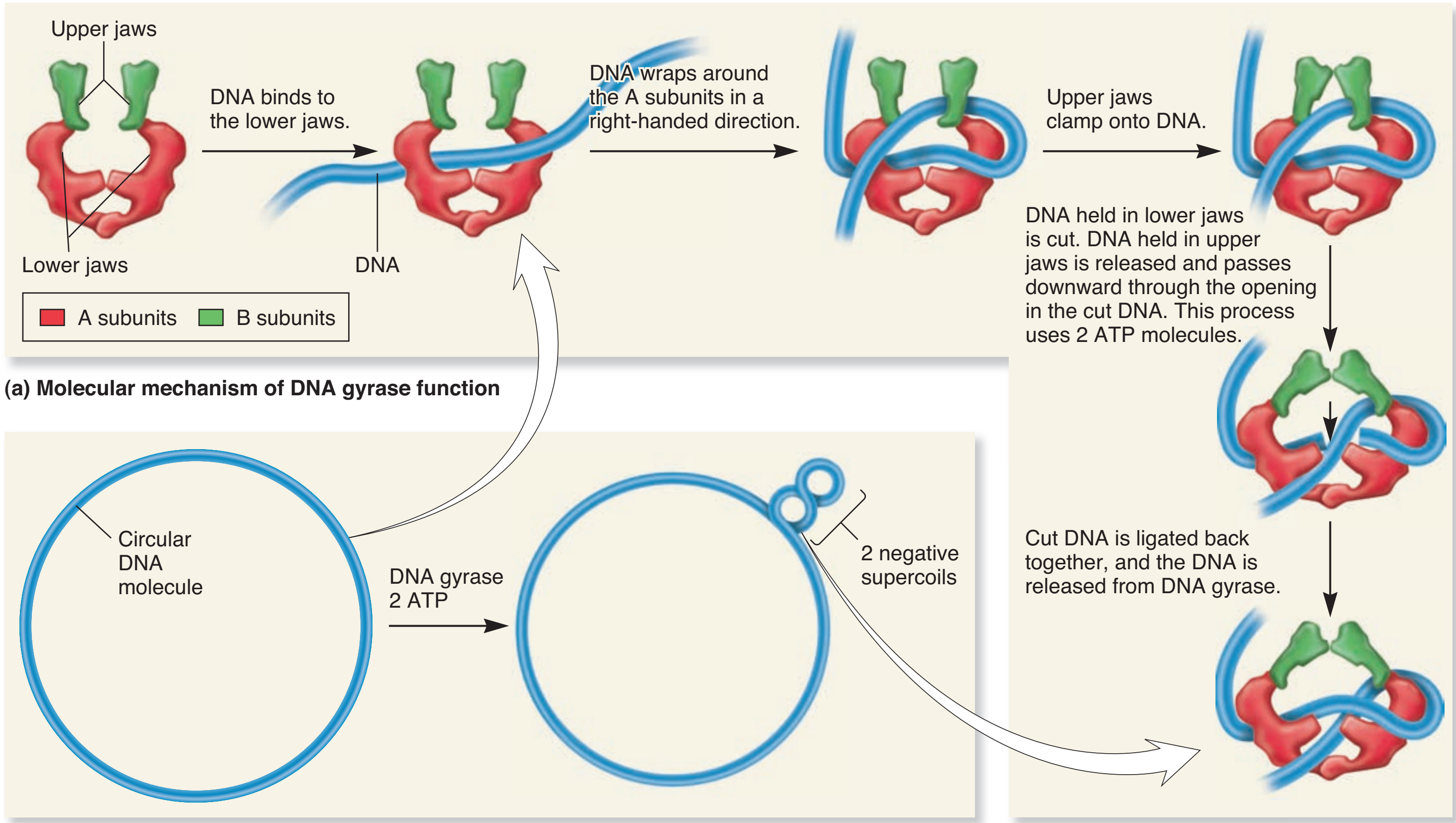
FIGURE 12.5 **Negative supercoiling promotes strand separation.**

Concept check: Why is strand separation sometimes beneficial?

jaws and then is wrapped in a right-handed direction around the two A subunits. The upper jaws then clamp onto another region of DNA. The DNA in the lower jaws is cut in both strands, and the other region of DNA is then released from the upper jaws and passed through this double-stranded break. The net result is that two negative supercoils have been introduced into the DNA molecule (**Figure 12.6b**). In addition, DNA gyrase in bacteria and topoisomerase II in eukaryotes can untangle DNA molecules. For example, as discussed in Chapter 13, circular DNA molecules are sometimes intertwined following DNA replication (see Figure 13.14). Such interlocked molecules can be separated via topoisomerase II.

A second type of enzyme, **topoisomerase I,** can relax negative supercoils. This enzyme can bind to a negatively supercoiled region and introduce a break in one of the DNA strands. After one DNA strand has been broken, the DNA molecule can rotate to relieve the tension that is caused by negative supercoiling. The broken strand is then resealed. The competing actions of DNA gyrase and topoisomerase I govern the overall supercoiling of the bacterial DNA.

The ability of gyrase to introduce negative supercoils into DNA is critical for bacteria to survive. For this reason, much research has been aimed at identifying drugs that specifically block the function of bacterial gyrase as a way to cure or alleviate diseases caused by bacteria. Two main classes—quinolones and coumarins—inhibit bacterial topoisomerases, thereby


Upper jaws
Lower jaws
DNA binds to the lower jaws.
DNA
DNA wraps around the A subunits in a right-handed direction.
Upper jaws clamp onto DNA.
🟥 A subunits
🟩 B subunits
DNA held in lower jaws is cut. DNA held in upper jaws is released and passes downward through the opening in the cut DNA. This process uses 2 ATP molecules.
Cut DNA is ligated back together, and the DNA is released from DNA gyrase.
Circular DNA molecule
DNA gyrase
2 ATP
2 negative supercoils

(a) Molecular mechanism of DNA gyrase function

(b) Overview of DNA gyrase function

FIGURE 12.6 **The action of DNA gyrase.** (**a**) DNA gyrase, also known as topoisomerase II, is composed of two A and two B subunits. The upper and lower jaws clamp onto two regions of DNA. The lower region is wrapped around the A subunits, which then cleave this DNA. The unbroken segment of DNA is released from the upper jaws and passes through the break. The break is repaired. The B subunits capture the energy from ATP hydrolysis to catalyze this process. The result is that two negative turns have been introduced into the DNA molecule. (**b**) Overview of DNA gyrase function.

Concept check: *In your own words, describe the step that requires the use of ATP.*

blocking bacterial cell growth. These drugs do not inhibit eukaryotic topoisomerases, which are structurally different from their bacterial counterparts. This finding has been the basis for the production of many drugs with important antibacterial applications. An example is ciprofloxacin (known also by the brand name Cipro), which is used to treat a wide spectrum of bacterial diseases, including anthrax.

12.2 REVIEWING THE KEY CONCEPTS

- Bacterial chromosomes are found in the nucleoid region of a bacterial cell (see Figure 12.2).
- Bacterial chromosomes are made more compact by the formation of loop domains and DNA supercoiling (see Figures 12.3, 12.4).
- Negative DNA supercoiling can promote DNA strand separation (see Figure 12.5).
- DNA gyrase (topoisomerase II) is a bacterial enzyme that introduces negative supercoils (see Figure 12.6).
- Topoisomerase I relaxes negative supercoils.

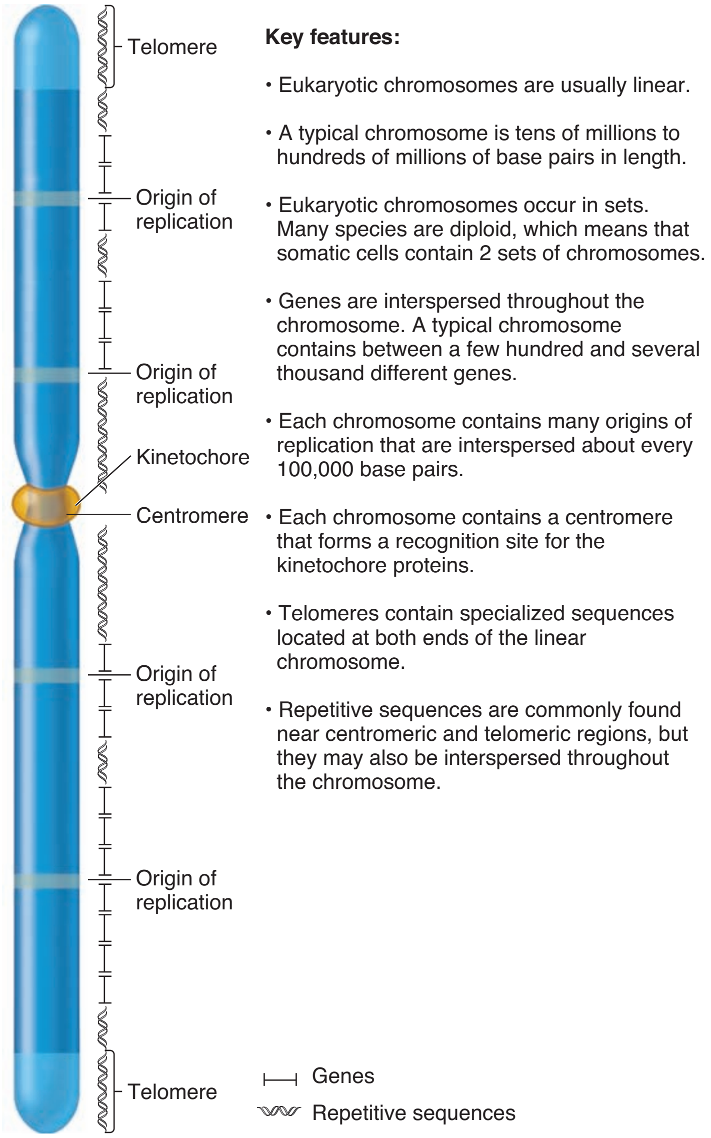
12.2 COMPREHENSION QUESTIONS

1. Mechanisms that make the bacterial chromosome more compact include
    a. the formation of loop domains.
    b. DNA supercoiling.
    c. crossing over.
    d. both a and b.
2. Negative supercoiling may enhance activities like transcription and DNA replication because it
    a. allows the binding of proteins to the major groove.
    b. promotes DNA strand separation.
    c. makes the DNA more compact.
    d. all of the above.
3. DNA gyrase
    a. promotes negative supercoiling.
    b. relaxes positive supercoils.
    c. cuts DNA strands as part of its function.
    d. does all of the above.

12.3 ORGANIZATION OF SITES ALONG EUKARYOTIC CHROMOSOMES

Eukaryotic species have one or more sets of chromosomes in the cell nucleus; each set is composed of several different linear chromosomes (refer back to Figure 8.1). Humans, for example, have two sets of 23 chromosomes each, for a total of 46. Each eukaryotic chromosome contains a long, linear DNA molecule (**Figure 12.7**). Three types of regions are required for chromosomal replication and segregation: origins of replication, centromeres, and telomeres. As mentioned previously, origins of replication are chromosomal sites that are necessary to initiate DNA replication. Unlike most bacterial chromosomes, which contain only one origin of replication, eukaryotic chromosomes contain many origins, interspersed approximately every 100,000 bp. The function of origins of replication is discussed in greater detail in Chapter 13.

Centromeres are regions that play a role in the proper segregation of chromosomes during mitosis and meiosis. For most species, each eukaryotic chromosome contains a single centromere, which usually appears as a constricted region of a mitotic chromosome. Centromeres function as a site for the formation of kinetochores, which assemble just before and during the very early stages of mitosis and meiosis. The **kinetochore** is composed of a group of cellular proteins that link the centromere to the spindle apparatus during mitosis and meiosis, ensuring the proper segregation of the chromosomes to each daughter cell. In certain yeast species, such as *Saccharomyces cerevisiae*, centromeres have a defined DNA sequence that is about 125 bp in length. This type of centromere is called a point centromere. By comparison, the centromeres found in more complex eukaryotes are much larger and contain many copies of tandemly repeated DNA sequences. (Tandem repeats are discussed later in this chapter.) These are called regional centromeres. They can range in size from several thousand base pairs in length to over one million. The repeated DNA sequences within regional centromeres by themselves are not necessary or sufficient to form a functional centromere with a kinetochore. Instead, other biochemical properties are needed to make a functional centromere. For example, a distinctive feature of all eukaryotic centromeres is that histone H3, which is discussed later in this chapter, is replaced with a histone variant called CENP-A. (Histone variants are described in Chapter 17.) However, researchers are still trying to identify all of the biochemical properties that distinguish regional centromeres and understand how these properties are transmitted during cell division.



Key features:

- Eukaryotic chromosomes are usually linear.
- A typical chromosome is tens of millions to hundreds of millions of base pairs in length.
- Eukaryotic chromosomes occur in sets. Many species are diploid, which means that somatic cells contain 2 sets of chromosomes.
- Genes are interspersed throughout the chromosome. A typical chromosome contains between a few hundred and several thousand different genes.
- Each chromosome contains many origins of replication that are interspersed about every 100,000 base pairs.
- Each chromosome contains a centromere that forms a recognition site for the kinetochore proteins.
- Telomeres contain specialized sequences located at both ends of the linear chromosome.
- Repetitive sequences are commonly found near centromeric and telomeric regions, but they may also be interspersed throughout the chromosome.

FIGURE 12.7 **Organization of eukaryotic chromosomes.** This organization refers to chromosomes in the cell nucleus. Mitochondrial and chloroplast chromosomes are described in Chapter 6.

Concept check: What are some differences between the types of sequences found in eukaryotic chromosomes versus bacterial chromosomes?

At the ends of linear chromosomes are found specialized regions known as **telomeres.** Telomeres serve several important functions in the replication and stability of the chromosome. As discussed in Chapter 8, telomeres prevent chromosomal rearrangements such as translocations. In addition, they prevent chromosome shortening in two ways. First, the telomeres protect chromosomes from digestion via enzymes called exonucleases that recognize the ends of DNA. Second, an unusual form of DNA replication may occur at the telomere to ensure that eukaryotic chromosomes do not become shortened with each round of DNA replication (see Chapter 13).

Genes are located between the centromeric and telomeric regions along the entire eukaryotic chromosome. A single chromosome usually has a few hundred to several thousand different genes. The sequence of a typical eukaryotic gene is several thousand to tens of thousands of base pairs in length. In less complex eukaryotes such as yeast, genes are relatively small and primarily contain nucleotide sequences that encode the amino acid sequences within proteins. In more complex eukaryotes such as mammals and flowering plants, structural genes tend to be much longer due to the presence of noncoding intervening sequences called **introns.** Introns range in size from less than 100 bp to more than 10,000 bp. Therefore, the presence of large introns can greatly increase the lengths of eukaryotic genes.

12.3 REVIEWING THE KEY CONCEPTS

- Eukaryotic chromosomes are usually linear and contain multiple origins of replication, a centromere, telomeres, repetitive sequences, and many genes (see Figure 12.7).

12.3 COMPREHENSION QUESTIONS

1. The chromosomes of eukaryotes typically contain
   a. a few hundred to several thousand different genes.
   b. multiple origins of replication.
   c. a centromere.
   d. telomeres at their ends.
   e. all of the above.

12.4 SIZES OF EUKARYOTIC GENOMES AND REPETITIVE SEQUENCES

The total amount of DNA in cells of eukaryotic species is usually much greater than that in bacterial cells. This enables eukaryotic genomes to contain many more genes than their bacterial counterparts. In this section, we will examine the sizes of eukaryotic genomes and consider how repetitive sequences may greatly contribute to their overall size.

The Sizes of Eukaryotic Genomes Vary Substantially

Different eukaryotic species vary dramatically in the size of their genomes (**Figure 12.8a**; note that this is a log scale). In general, more complex species have larger genome sizes. For example, the genome size of a mammal is much larger than that of a yeast. However, in many cases, variation in genome size is not entirely related to the complexity of the species. As shown in **Figure 12.8b**, and **c**, two closely related species of salamander, *Plethodon richmondi* and *Plethodon larselli*, differ considerably in genome size. The genome of *P. larselli* is more than twice as large as the genome of *P. richmondi*. However, the genome of *P. larselli* probably doesn't contain more genes. How do we explain the difference in genome size? The additional DNA in *P. larselli* is due to the accumulation of repetitive DNA sequences present in many copies. In some species, these repetitive sequences can accumulate to enormous levels. Such highly repetitive sequences do not encode proteins, and their function remains a matter of controversy and great interest. The structure and significance of repetitive DNA is discussed next.

The Genomes of Eukaryotes Contain Sequences That Are Unique, Moderately Repetitive, or Highly Repetitive

The term **sequence complexity** refers to the number of times a particular base sequence appears throughout the genome. Unique or nonrepetitive sequences are those found once or a few times within the genome. Structural genes are typically unique sequences of DNA. The vast majority of proteins in eukaryotic cells are encoded by genes present in one or a few copies. In the case of humans, unique sequences make up roughly 41% of the entire genome (**Figure 12.9**). These include exons, which encode polypeptide sequences, and unique noncoding DNA such as sequences of introns and other chromosomal regions.

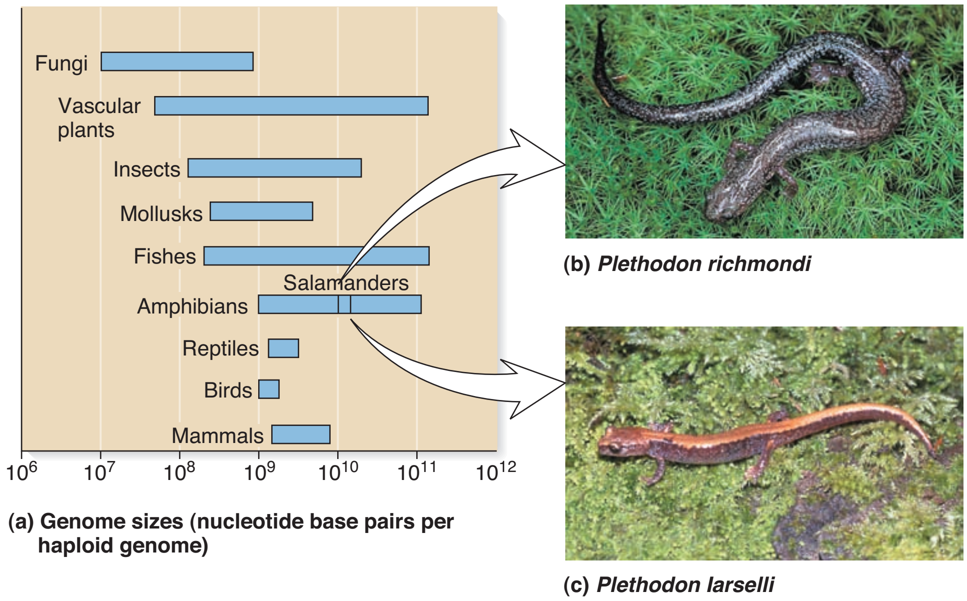
Moderately repetitive sequences are found a few hundred to several thousand times in the genome. In a few cases, moderately repetitive sequences are multiple copies of the same gene. For example, the genes that encode ribosomal RNA (rRNA) are found in many copies. Ribosomal RNA is necessary for the

FIGURE 12.8 **Haploid genome sizes among groups of eukaryotic species.** **(a)** Ranges of genome sizes among different groups of eukaryotes. **(b)** A species of salamander, *Plethodon richmondi,* and **(c)** a close relative, *Plethodon larselli.* The genome of *P. larselli* is over twice as large as that of *P. richmondi.*

Data from T. Ryan Gregory, J.A. Nichol, H. Tamm, et al. (2007) Eukaryotic genome size databases. *Nucleic Acids Res.* 35:D332–D338.

Genes→Traits The two species of salamander shown here have very similar traits, even though the genome of *P. larselli* is over twice as large as that of *P. richmondi*. However, the genome of *P. larselli* is not likely to contain more genes. Rather, the additional DNA is due to the accumulation of short repetitive DNA sequences that do not code for genes and are present in many copies.

Concept check: What are two reasons for the wide variation in genome sizes among eukaryotic species?



(a) Genome sizes (nucleotide base pairs per haploid genome)

(b) *Plethodon richmondi*

(c) *Plethodon larselli*

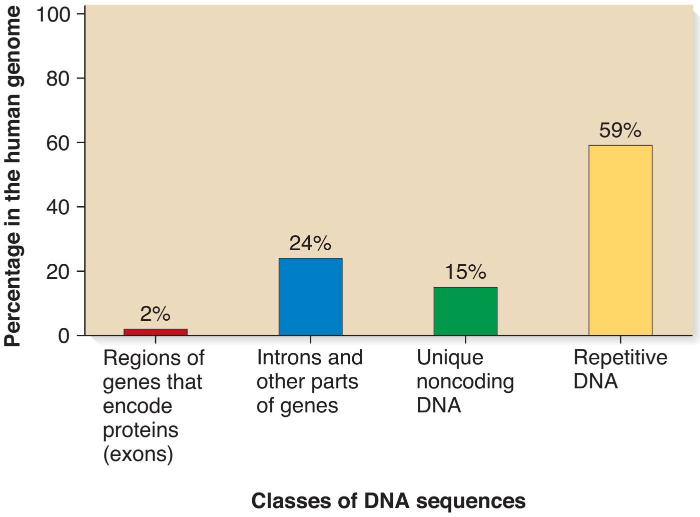


FIGURE 12.9 **Relative amounts of unique and repetitive DNA sequences in the human genome.**

functioning of ribosomes. Cells need a large amount of rRNA for making ribosomes, and this is accomplished by having multiple copies of the genes that encode rRNA. Likewise, the histone genes are also found in multiple copies because a large number of histone proteins are needed for the structure of eukaryotic chromosomes. In addition, other types of functionally important sequences can be moderately repetitive. For example, moderately repetitive sequences may play a role in the regulation of gene transcription and translation. By comparison, some moderately repetitive sequences do not play a functional role and are derived from **transposable elements (TE)**—segments of DNA that have the ability to move within a genome. This category of repetitive sequences is discussed in greater detail in Chapter 21.

Highly repetitive sequences are found tens of thousands or even millions of times throughout a genome. Each copy of a highly repetitive sequence is relatively short, ranging from a few nucleotides to several hundred in length. A widely studied example is the *Alu* family of sequences found in humans and other primates. The *Alu* sequence is approximately 300 bp long. This sequence derives its name from the observation that it contains a site for cleavage by a restriction enzyme known as *Alu*I. (The function of restriction enzymes is described in Chapter 19.) The *Alu* sequence is present in about 1,000,000 copies in the human genome. It represents about 10% of the total human DNA and occurs approximately every 5000 to 6000 bp! Evolutionary studies suggest that the *Alu* sequence arose 65 million years ago from a section of a single ancestral gene known as the *7SL RNA* gene. Since that time, this gene has become a type of TE called a **retroelement,** which can be transcribed into RNA, copied into DNA, and inserted into the genome. Remarkably, over the course of 65 million years, the *Alu* sequence has been copied and inserted into the human genome to achieve the modern number of about 1,000,000 copies.

Repetitive sequences, like the *Alu* family, are interspersed throughout the genome. However, some moderately and highly repetitive sequences are clustered together in a **tandem array,** also known as tandem repeats. In a tandem array, a very short nucleotide sequence is repeated many times in a row. In *Drosophila,* for example, 19% of the chromosomal DNA is highly repetitive DNA found in tandem arrays. An example is shown here.

AATATAATATAATATAATATAATATATAATAT
TTATATTATATTATATTATATTATATATTATA

In this particular tandem array, two related sequences, AATAT and AATATAT, are repeated. As mentioned, tandem arrays of short sequences are commonly found in centromeric regions of chromosomes and can be quite long, sometimes more than 1,000,000 bp in length!

What is the functional significance of highly repetitive sequences? Whether highly repetitive sequences have any significant function is controversial. Some experiments in *Drosophila* indicate that highly repetitive sequences may be important in the proper segregation of chromosomes during meiosis. It is not yet clear if highly repetitive DNA plays the same role in other species. The sequences within highly repetitive DNA vary greatly from species to species. Likewise, the amount of highly repetitive DNA can vary a great deal even among closely related species (as noted in the discussion of Figure 12.8a).

12.4 REVIEWING THE KEY CONCEPTS

- The genome sizes of eukaryotes vary greatly. Some of this variation is due to the accumulation of repetitive sequences (see Figure 12.8).
- The human genome contains about 41% unique sequences and 59% repetitive sequences (see Figure 12.9).

12.4 COMPREHENSION QUESTIONS

1. Which of the following is an example of a moderately repetitive sequence?
   a. rRNA genes
   b. Most structural genes
   c. Both a and b
   d. None of the above

12.5 STRUCTURE OF EUKARYOTIC CHROMOSOMES IN NONDIVIDING CELLS

A distinguishing feature of eukaryotic cells is that their chromosomes are located within a separate cellular compartment known as the **nucleus.** The DNA within a typical eukaryotic chromosome is a single, linear double-stranded molecule that may be hundreds of millions of base pairs in length. If the DNA from a single set of human chromosomes was stretched from end to end, the length would be over 1 meter! By comparison, most eukaryotic cells are only 10 to 100 μm in diameter, and the cell nucleus is only about 2 to 4 μm in diameter. Therefore, the DNA in a eukaryotic cell must be folded and packaged by a staggering amount to fit inside the nucleus. In eukaryotic chromosomes, as in bacterial chromosomes, this is accomplished by the binding of the DNA to many different cellular proteins. The DNA–protein complex found within eukaryotic chromosomes is termed **chromatin.** In recent years, it has become increasingly clear that the proteins bound to chromosomal DNA are subject to change over the life of the cell. These changes in protein composition, in turn, affect the degree of compaction of the chromatin. In this section, we consider how chromosomes are compacted during interphase—the period of the cell cycle that includes the G_1, S, and G_2 phases. In the following section, we examine the additional compaction that is necessary to produce the highly condensed chromosomes found in M phase.

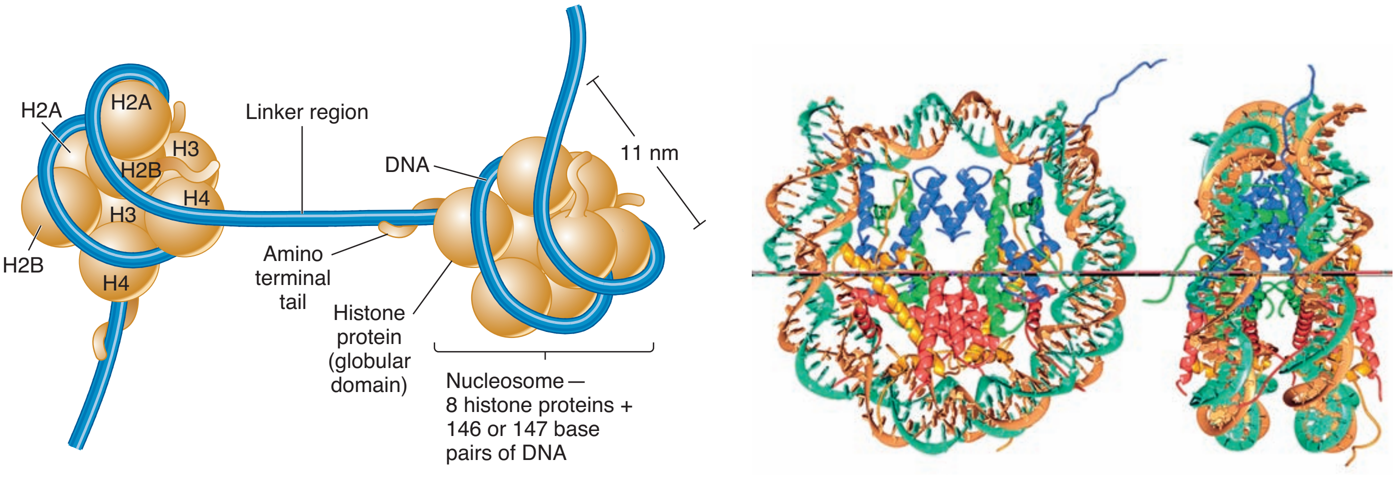
Linear DNA Wraps Around Histone Proteins to Form Nucleosomes, the Repeating Structural Unit of Chromatin

The repeating structural unit within eukaryotic chromatin is the **nucleosome**—a double-stranded segment of DNA wrapped around an octamer of **histone proteins** (**Figure 12.10a**). Each octamer contains eight histone subunits, two copies each of four different histone proteins. The DNA lies on the surface and makes 1.65 negative superhelical turns around the histone octamer. The amount of DNA required to wrap around the histone octamer is 146 or 147 bp. At its widest point, a single nucleosome is about 11 nm in diameter.

The chromatin of eukaryotic cells contains a repeating pattern in which the nucleosomes are connected by linker regions of DNA that vary in length from 20 to 100 bp, depending on the species and cell type. It has been suggested that the overall structure of connected nucleosomes resembles beads on a string. This structure shortens the length of the DNA molecule about sevenfold.

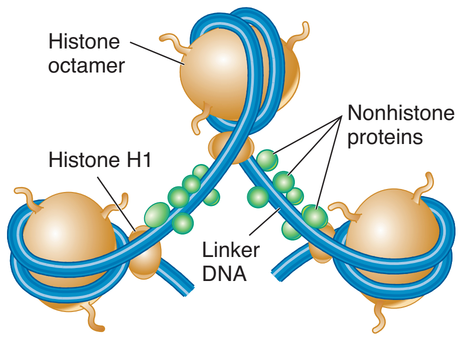
Each of the histone proteins consists of a globular domain and a flexible, charged amino terminus called an amino terminal tail. Histone proteins are very basic proteins because they contain a large number of positively charged lysine and arginine amino acids. The arginines, in particular, play a major role in binding to the DNA. Arginines within the histone proteins form electrostatic and hydrogen-bonding interactions with the phosphate groups along the DNA backbone. The octamer of histones contains two molecules each of four different histone proteins: H2A, H2B, H3, and H4. These are called the core histones. In 1997, Timothy Richmond and colleagues determined the structure of a nucleosome by X-ray crystallography (**Figure 12.10b**).

Another histone, H1, is found in most eukaryotic cells and is called the linker histone. It binds to the DNA in the linker region between nucleosomes and may help to compact adjacent nucleosomes (**Figure 12.10c**). The linker histones are less tightly


H2A
H2A
H3
H2B
H4
H3
H2B
H4
Linker region
DNA
11 nm
Amino terminal tail
Histone protein (globular domain)
Nucleosome — 8 histone proteins + 146 or 147 base pairs of DNA

(a) Nucleosomes showing core histone proteins

(b) Molecular model for nucleosome structure



(c) Nucleosomes showing linker histones and nonhistone proteins

FIGURE 12.10 **Nucleosome structure.** **(a)** A nucleosome consists of 146 or 147 bp of DNA wrapped around an octamer of core histone proteins. **(b)** A model for the structure of a nucleosome as determined by X-ray crystallography. This drawing shows two views of a nucleosome that are at right angles to each other. The horizontal bar passes through the center of the nucleosome in each view. **(c)** The linker region of DNA connects adjacent nucleosomes. The linker histone H1 and nonhistone proteins also bind to this linker region.

(Image courtesy of Timothy J. Richmond. Reprinted by permission from Macmillan Publishers Ltd. Luger K., Mader, AW, Richmond, RK, Sargent, DF, Richmond, TJ [1997] Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389:6648, 251–260.)

Concept check: What is the diameter of a nucleosome?

bound to the DNA than are the core histones. In addition, nonhistone proteins bound to the linker region play a role in the organization and compaction of chromosomes, and their presence may affect the expression of nearby genes.

The Repeating Nucleosome Structure Is Revealed by Digestion of the Linker Region

The model of nucleosome structure was originally proposed by Roger Kornberg in 1974. He based his proposal on several observations. Biochemical experiments had shown that chromatin contains a ratio of one molecule of each of the four core histones (namely, H2A, H2B, H3, and H4) per 100 bp of DNA. Approximately one H1 protein was found per 200 bp of DNA. In addition, purified core histone proteins were observed to bind to each other via specific pairwise interactions. Subsequent X-ray diffraction studies showed that chromatin is composed of a repeating pattern of smaller units. Finally, electron microscopy of chromatin fibers revealed a diameter of approximately 11 nm. Taken together, these observations led Kornberg to propose a model in which the DNA double helix is wrapped around an octamer of core histone proteins. Including the linker region, this involves about 200 bp of DNA.

Markus Noll decided to test Kornberg's model by digesting chromatin with DNase I, an enzyme that cuts the DNA backbone. He then accurately measured the molecular mass of the DNA fragments by gel electrophoresis. Noll assumed that the linker region of DNA is more accessible to DNase I and, therefore, DNase I is more likely to make cuts in the linker region than in the 146-bp region that is tightly bound to the core histones. If this is correct, incubation with DNase I is expected to make cuts in the linker region, thereby producing DNA pieces approximately 200 bp in length. The size of the DNA fragments may vary somewhat because the linker region is not of constant length and because the cut within the linker region may occur at different sites.

Figure 12.11 describes Noll's experimental protocol. He began with nuclei from rat liver cells and incubated them with low, medium, or high concentrations of DNase I. The DNA was extracted into an aqueous phase and then loaded onto an agarose gel that separated the fragments according to their molecular mass. The DNA fragments within the gel were stained with a UV-sensitive dye, ethidium bromide, which made it possible to view the DNA fragments under UV illumination.

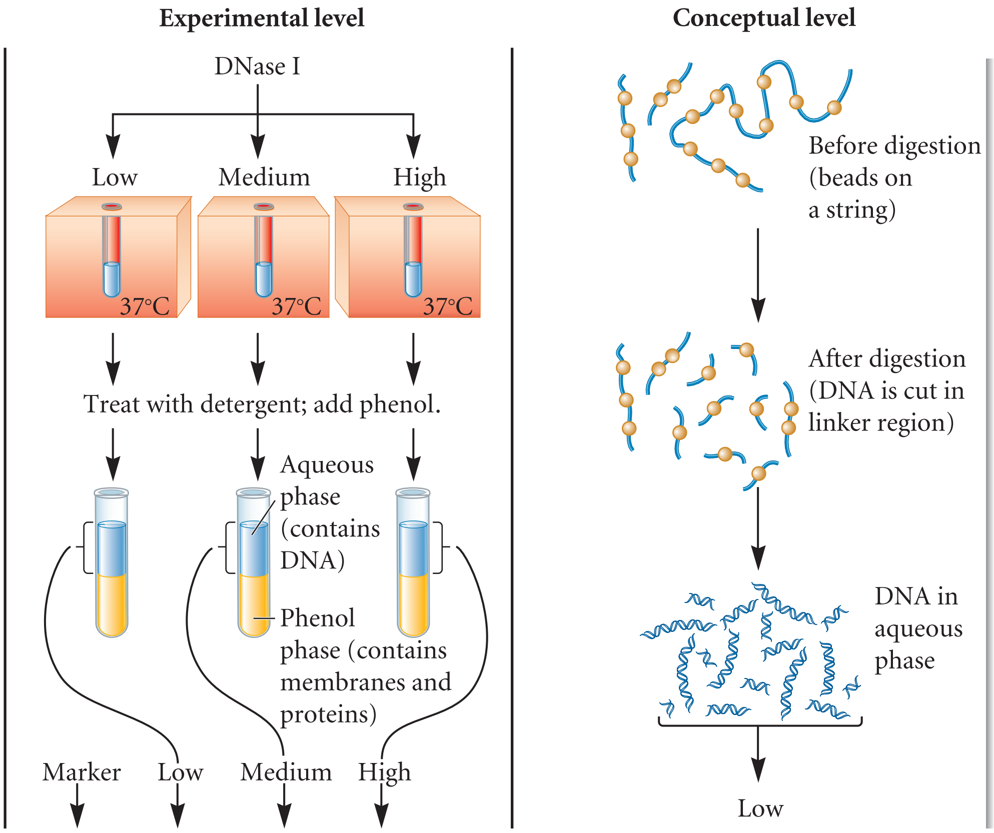
▶ THE HYPOTHESIS

The aim of this experiment was to test the beads-on-a-string model for chromatin structure. According to this model, DNase I should preferentially cut the DNA in the linker region, thereby producing DNA pieces that are about 200 bp in length.

▶ TESTING THE HYPOTHESIS — FIGURE 12.11 **DNase I cuts chromatin into repeating units containing 200 bp of DNA.**

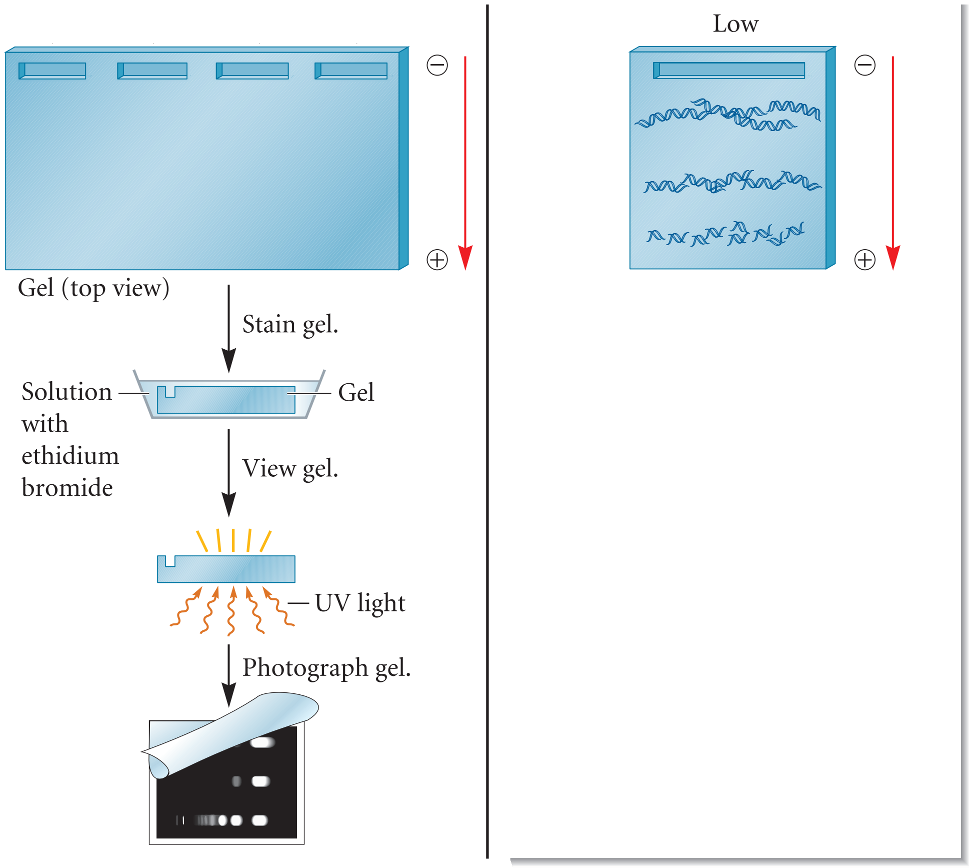
Starting material: Nuclei from rat liver cells.

1. Incubate the nuclei with low, medium, and high concentrations of DNase I. The conceptual level illustrates a low DNase I concentration.
2. Extract the DNA. This involves dissolving the nuclear membrane with detergent and extracting with the organic solvent phenol.

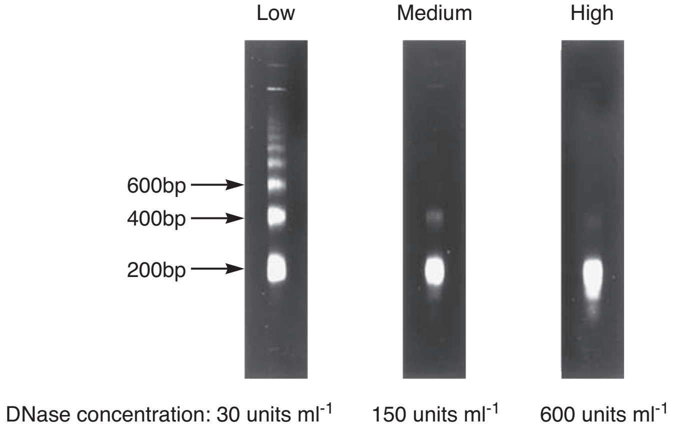

Experimental level
DNase I
Low
Medium
High
37°C
37°C
37°C
Treat with detergent; add phenol.
Aqueous phase (contains DNA)
Phenol phase (contains membranes and proteins)
Marker
Low
Medium
High
Conceptual level
Before digestion (beads on a string)
After digestion (DNA is cut in linker region)
DNA in aqueous phase
Low

3. Load the DNA into a well of an agarose gel and run the gel to separate the DNA pieces according to size. On this gel, also load DNA fragments of known molecular mass (marker lane).

4. Visualize the DNA fragments by staining the DNA with ethidium bromide, a dye that binds to DNA and is fluorescent when excited by UV light.



▶ THE DATA



(Reprinted by permission from Macmillan Publishers Ltd. Noll M [1974] Subunit structure of chromatin. *Nature. 251*:249–251.)

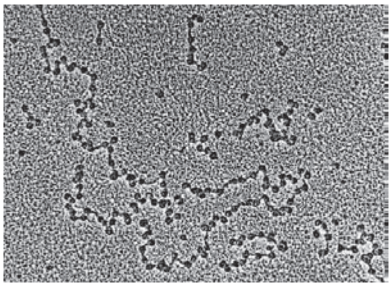
▶ INTERPRETING THE DATA

As shown in the data of Figure 12.11, at a high DNase I concentration, the entire sample of chromosomal DNA was digested into fragments of approximately 200 bp in length. This result is predicted by the beads-on-a-string model. Furthermore, at a low or medium DNase I concentration, longer pieces were observed, and these were in multiples of 200 bp (400, 600, etc.). How do we explain these longer pieces? They occurred because occasional linker regions remained uncut at a low or medium DNase I concentration. For example, if one linker region was not cut, a DNA piece would contain two nucleosomes and be 400 bp in length. If two consecutive linker regions were not cut, a piece with three nucleosomes containing about 600 bp of DNA would result. Taken together, these results strongly supported the nucleosome model for chromatin structure.

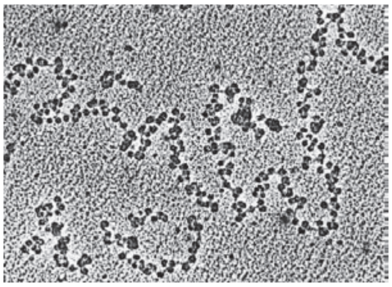
A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

Nucleosomes Become Closely Associated to Form a 30-nm Fiber

In eukaryotic chromatin, nucleosomes associate with each other to form a more compact structure that is 30 nm in diameter. Evidence for the packaging of nucleosomes was obtained in the microscopy studies of Fritz Thoma in 1977. Chromatin samples were treated with a resin that removed histone H1, but the removal depended on the salt concentration. A moderate salt solution (100 mM NaCl) removed H1, but a solution with no added NaCl did not remove H1. These samples were then observed with an electron microscope. At moderate salt concentrations (**Figure 12.12a**), the chromatin exhibited the classic beads-on-a-string morphology. Without added NaCl (when H1 is expected to remain bound to the DNA), these "beads" associated with each other into a more compact conformation (**Figure 12.12b**). These results suggest that the nucleosomes are packaged



(a) H1 histone not bound—beads on a string



(b) H1 histone bound to linker region—nucleosomes more compact

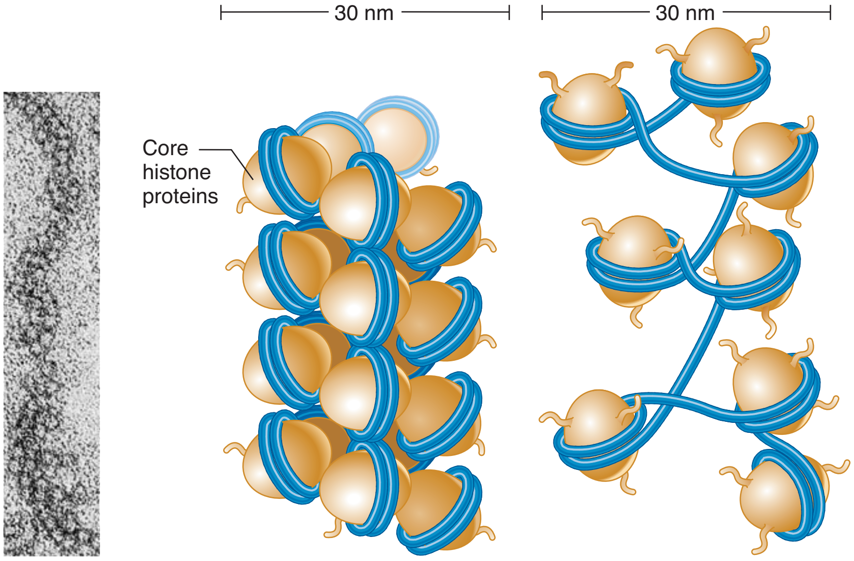
FIGURE 12.12 **The nucleosome structure of eukaryotic chromatin as viewed by electron microscopy.** The chromatin in **(a)** has been treated with a moderate salt concentration to remove the linker histone H1. It exhibits the classic beads-on-a-string morphology. The chromatin in **(b)** has been incubated without added NaCl and shows a more compact morphology.

into a more compact unit and that H1 has a role in the packaging of nucleosomes. However, the precise role of H1 in chromatin compaction remains unclear. Recent data suggest that the core histones also play a key role in the compaction and relaxation of chromatin.

The experiment of Figure 12.12 and other experiments established that nucleosome units are organized into a more compact structure that is 30 nm in diameter, known as the **30-nm fiber** (**Figure 12.13a**). The 30-nm fiber shortens the total length of DNA another sevenfold. The structure of the 30-nm fiber has proven difficult to determine, because the conformation of the DNA may be substantially altered when it is extracted from living cells. Most models for the 30-nm fiber fall into two main classes. The solenoid model suggests a helical structure in which contact between nucleosomes produces a symmetrically compact structure within the 30-nm fiber (**Figure 12.13b**). This type of model is still favored by some researchers in the field. However, experimental data also suggest that the 30-nm fiber may not form such a regular structure. Instead, an alternative zigzag model, advocated by Rachel Horowitz, Christopher Woodcock, and others, is based on techniques such as cryoelectron microscopy (electron microscopy at low temperature). According to the zigzag model, linker regions within the 30-nm structure are variably bent and twisted, and little face-to-face contact occurs between nucleosomes (**Figure 12.13c**). At this level of compaction, the overall picture of chromatin that emerges is an irregular, fluctuating, three-dimensional zigzag structure with stable nucleosome units connected by deformable linker regions. In 2005, Timothy Richmond and colleagues were the first to solve the crystal structure of a segment of DNA containing four nucleosomes. The structure revealed that the linker region of DNA zigzags back and forth between each nucleosome, a feature consistent with the zigzag model.

Chromosomes Are Further Compacted by Anchoring the 30-nm Fiber into Radial Loop Domains Along the Nuclear Matrix

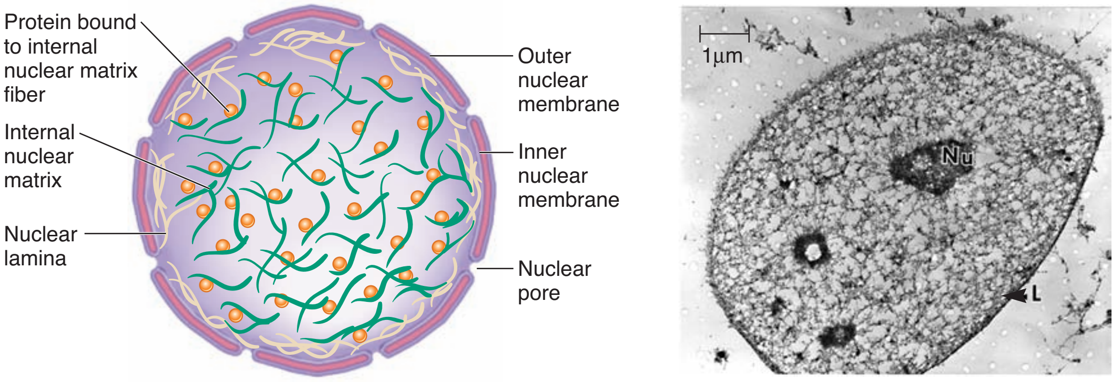
Thus far, we have examined two mechanisms that compact eukaryotic DNA. These involve the wrapping of DNA within nucleosomes and the arrangement of nucleosomes to form a 30-nm fiber. Taken together, these two events shorten the DNA nearly 50-fold. A third level of compaction involves interactions



(a) Micrograph of a 30 nm fiber **(b) Solenoid model** **(c) Zigzag model**

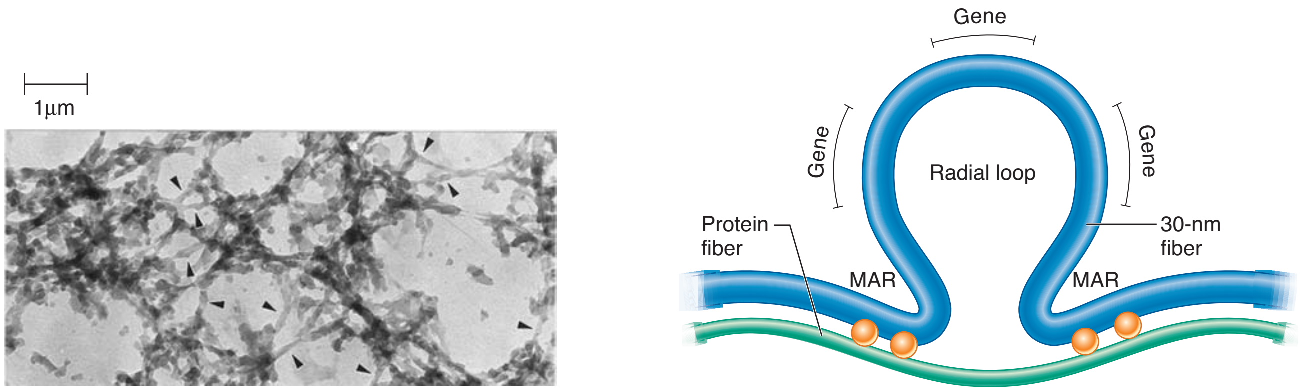
FIGURE 12.13 **The 30-nm fiber.** **(a)** A photomicrograph of the 30-nm fiber. **(b)** In the solenoid model, the nucleosomes are packed in a spiral configuration. **(c)** In the zigzag model, the linker DNA forms a more irregular structure, and less contact occurs between adjacent nucleosomes. The zigzag model is consistent with more recent data regarding chromatin conformation.

Concept check: Describe the distinguishing features of the solenoid model and zigzag model.



(a) Proteins that form the nuclear matrix

(b) Micrograph of nucleus with chromatin removed



(c) Micrograph showing a close-up of nuclear matrix

(d) Radial loop bound to a nuclear matrix fiber

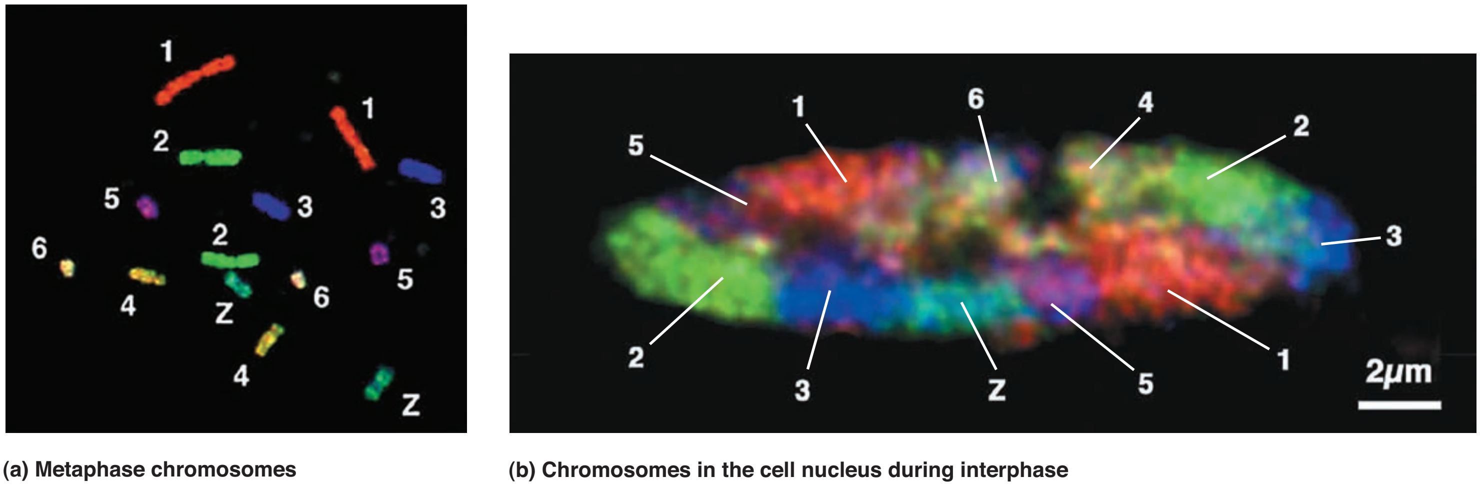
FIGURE 12.14 **Structure of the nuclear matrix.** **(a)** This schematic drawing shows the arrangement of the matrix within a cell nucleus. The nuclear lamina, which lines the inner nuclear membrane, is a collection of fibrous proteins that line the inner nuclear membrane. The internal nuclear matrix is composed of protein filaments (depicted in green) that may be interconnected. These fibers also have many other proteins associated with them (depicted in orange). **(b)** An electron micrograph of the nuclear matrix during interphase after the chromatin has been removed. The nucleolus is labeled Nu, and the lamina is labeled L. **(c)** At higher magnification, the protein fibers are more easily seen (arrowheads point at fibers). **(d)** The matrix-attachment regions (MARs), which contain a high percentage of A and T bases, bind to the nuclear matrix and create radial loops. This causes a greater compaction of eukaryotic chromosomal DNA.

Concept check: *What is the function of the nuclear matrix?*

between the 30-nm fibers and a filamentous network of proteins in the nucleus called the **nuclear matrix.** As shown in Figure 12.14a, the nuclear matrix consists of two parts. The **nuclear lamina** is a collection of fibers that line the inner nuclear membrane. These fibers are composed of intermediate filament proteins. The second part is an **internal nuclear matrix,** which is connected to the nuclear lamina and fills the interior of the nucleus. The internal nuclear matrix, whose structure and functional role remain controversial, is hypothesized to be an intricate fine network of irregular protein fibers plus many other proteins that bind to these fibers. Even when the chromatin is extracted from the nucleus, the internal nuclear matrix may remain intact (Figure 12.14b and c). However, the matrix should not be considered a static structure. Research indicates that the protein composition of the internal nuclear matrix is very dynamic and complex, consisting of dozens or perhaps hundreds of different proteins. The protein composition varies depending on species, cell type, and environmental conditions. This complexity has made it difficult to propose models regarding its overall organization. Further research is necessary to understand the structure and dynamic nature of the internal nuclear matrix.

The proteins of the nuclear matrix are involved in compacting the DNA into **radial loop domains,** similar to the loop domains described for the bacterial chromosome. During interphase, chromatin is organized into loops, often 25,000 to 200,000 bp in size, which are anchored to the nuclear matrix. The chromosomal DNA of eukaryotic species contains sequences called **matrix-attachment regions (MARs),** which are also called scaffold-attachment regions (SARs). They are interspersed at regular intervals throughout the genome. The MARs bind to specific proteins in the nuclear matrix, thus forming chromosomal loops (Figure 12.14d).

Why is the attachment of radial loops to the nuclear matrix important? In addition to compaction, the nuclear matrix serves to organize the chromosomes within the nucleus. Each chromosome in the cell nucleus is located in a discrete **chromosome territory.**



(a) Metaphase chromosomes

(b) Chromosomes in the cell nucleus during interphase

FIGURE 12.15 **Chromosome territories in the cell nucleus.** **(a)** Several metaphase chromosomes from the chicken were labeled with chromosome-specific probes. Each of seven types of chicken chromosomes (i.e., 1, 2, 3, 4, 5, 6, and Z) is labeled a different color. **(b)** The same probes were used to label interphase chromosomes in the cell nucleus. Each of these chromosomes occupies its own distinct, nonoverlapping territory within the cell nucleus. (*Note:* Chicken cells are diploid, with two copies of each chromosome. The Z chromosome is a sex chromosome.)

(Reprinted by permission from Macmillan Publishers Ltd. T. Cremer, C. Cremer. [2001] Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nature Rev Genet 2:4, 292–301.)

Concept check: What is a chromosome territory?

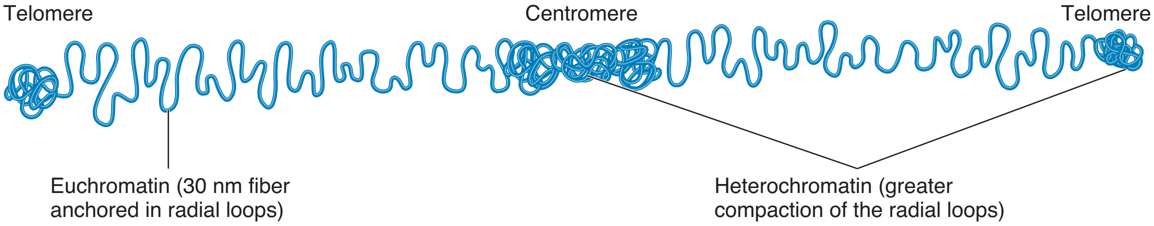

Telomere
Centromere
Telomere
Euchromatin (30 nm fiber anchored in radial loops)
Heterochromatin (greater compaction of the radial loops)

FIGURE 12.16 **Chromatin structure during interphase.** Heterochromatic regions are more highly condensed and tend to be localized in centromeric and telomeric regions.

Concept check: Would you expect to find active genes in heterochromatic or euchromatic regions?

As shown in studies by Thomas Cremer, Christoph Cremer, and others, these territories can be viewed when interphase cells are exposed to multiple fluorescent molecules that recognize specific sequences on particular chromosomes. Figure 12.15 illustrates an experiment in which chicken cells were exposed to a mixture of probes that recognize multiple sites along several of the larger chromosomes found in this species (*Gallus gallus*). Figure 12.15a shows the chromosomes in metaphase. The probes label each type of metaphase chromosome with a different color. Figure 12.15b shows the use of the same probes during interphase, when the chromosomes are less condensed and found in the cell nucleus. As seen here, each chromosome occupies its own distinct territory.

Before ending the topic of interphase chromosome compaction, let's consider how the compaction level of interphase chromosomes may vary. This variability can be seen with a light microscope and was first observed by the German cytologist Emil Heitz in 1928. He coined the term **heterochromatin** to describe the tightly compacted regions of chromosomes. In general, these regions of the chromosome are transcriptionally inactive. By comparison, the less condensed regions, known as **euchromatin,** are usually capable of gene transcription. In euchromatin, the 30-nm fiber forms radial loop domains. In heterochromatin, these radial loop domains become compacted even further.

Figure 12.16 illustrates the distribution of euchromatin and heterochromatin in a typical eukaryotic chromosome during interphase. The chromosome contains regions of both heterochromatin and euchromatin. Heterochromatin is most abundant in the centromeric regions of the chromosomes and, to a lesser extent, in the telomeric regions. The term **constitutive heterochromatin** refers to chromosomal regions that are always heterochromatic and permanently inactive with regard to transcription. Constitutive heterochromatin usually contains highly repetitive DNA sequences, such as tandem repeats, rather than gene sequences. **Facultative heterochromatin** refers to chromatin that can occasionally interconvert between heterochromatin and euchromatin. An example of facultative heterochromatin occurs in female mammals when one of the two X chromosomes is converted to a heterochromatic Barr body. As discussed in Chapter 4, most of the genes on the Barr body are transcriptionally inactive. The conversion of one X chromosome to heterochromatin occurs during embryonic development in the somatic cells of the body.

12.5 REVIEWING THE KEY CONCEPTS

- The term *chromatin* refers to the DNA–protein complex found within eukaryotic chromosomes.
- Eukaryotic DNA wraps around an octamer of histone proteins to form nucleosomes (see Figure 12.10).
- Noll tested Kornberg's nucleosome model by digesting eukaryotic chromatin with varying concentrations of DNase I (see Figure 12.11).
- The linker histone, H1, plays a role in nucleosome compaction (see Figure 12.12).
- Nucleosomes are further compacted to form a 30-nm fiber. Solenoid and zigzag models have been proposed (see Figure 12.13).
- Chromatin is further compacted by the attachment of 30-nm fibers to protein filaments to form radial loop domains (see Figure 12.14).
- Within the cell nucleus, each eukaryotic chromosome occupies its own unique chromosome territory (see Figure 12.15).
- In nondividing cells, each chromosome has highly compacted regions called heterochromatin and less compacted regions called euchromatin (see Figure 12.16).

12.5 COMPREHENSION QUESTIONS

1. What are the components of a single nucleosome?
   a. About 146 bp of DNA and four histone proteins
   b. About 146 bp of DNA and eight histone proteins
   c. About 200 bp of DNA and four histone proteins
   d. About 200 bp of DNA and eight histone proteins
2. In Noll's experiment to test the beads-on-a-string model, exposure of nuclei to a low concentration of DNase I resulted in
   a. a single band of DNA with a size of approximately 200 bp.
   b. several bands of DNA in multiples of 200 bp.
   c. a single band of DNA with a size of 100 bp.
   d. several bands of DNA in multiples of 100 bp.
3. With regard to the 30-nm fiber, a key difference between the solenoid and zigzag models is
   a. the solenoid model suggests a helical structure.
   b. the zigzag model suggests a more irregular pattern of nucleosomes.
   c. the zigzag model does not contain nucleosomes.
   d. both a and b are correct.
4. A chromosome territory is a region
   a. along a chromosome where many genes are clustered.
   b. along a chromosome where the nucleosomes are close together.
   c. in a cell nucleus where a single chromosome is located.
   d. in a cell nucleus where multiple chromosomes are located.

12.6 STRUCTURE OF EUKARYOTIC CHROMOSOMES DURING CELL DIVISION

As described in Chapter 2, when eukaryotic cells prepare to divide, the chromosomes become very condensed or compacted. This aids in their proper sorting during metaphase. The highly compacted chromosomes that are viewed during metaphase are called metaphase chromosomes. In this section, we will examine the compaction level of metaphase chromosomes and discuss certain proteins that are involved in the process.

Metaphase Chromosomes Are Highly Compacted

Figure 12.17 illustrates the levels of compaction that lead to a metaphase chromosome. During interphase, most of the chromosomal DNA is found in euchromatin, in which the 30-nm fibers form radial loop domains that are attached to a protein **scaffold** formed from nonhistone proteins of the nuclear matrix. The average distance that loops radiate from the protein scaffold is approximately 300 nm. This structure can be further compacted via additional folding of the radial loop domains and protein scaffold. This further level of compaction greatly shortens the overall length of a chromosome and produces a diameter of approximately 700 nm, which is the compaction level found in heterochromatin. During interphase, most chromosomal regions are euchromatic, and some localized regions, such as those near centromeres, are heterochromatic.

As cells enter M phase, the level of compaction changes dramatically. By the end of prophase, sister chromatids are entirely heterochromatic. Two parallel chromatids have a larger diameter of approximately 1400 nm but a much shorter length compared with interphase chromosomes. These highly condensed metaphase chromosomes undergo little gene transcription because it is difficult for transcription proteins to gain access to the compacted DNA. Therefore, most transcriptional activity ceases during M phase, although a few specific genes may be transcribed. M phase is usually a short period of the cell cycle.

In highly condensed chromosomes, such as those found in metaphase, the radial loops are highly compacted and remain anchored to the scaffold. **Figure 12.18a** shows a human metaphase chromosome. In this condition, the radial loops of DNA are in a very compact configuration. If this chromosome is treated with a high concentration of salt to remove both the core and linker histones, the highly compact configuration is lost, but the bottoms of the elongated loops remain attached to the scaffold composed of nonhistone proteins. In **Figure 12.18b**, an arrow points to an elongated DNA strand emanating from the darkly staining scaffold. Remarkably, the scaffold retains the shape of the original metaphase chromosome even though the DNA strands have become greatly elongated. These results illustrate that the structure of metaphase chromosomes is determined by the nuclear matrix proteins, which form the scaffold, and by the histones, which are needed to compact the radial loops.

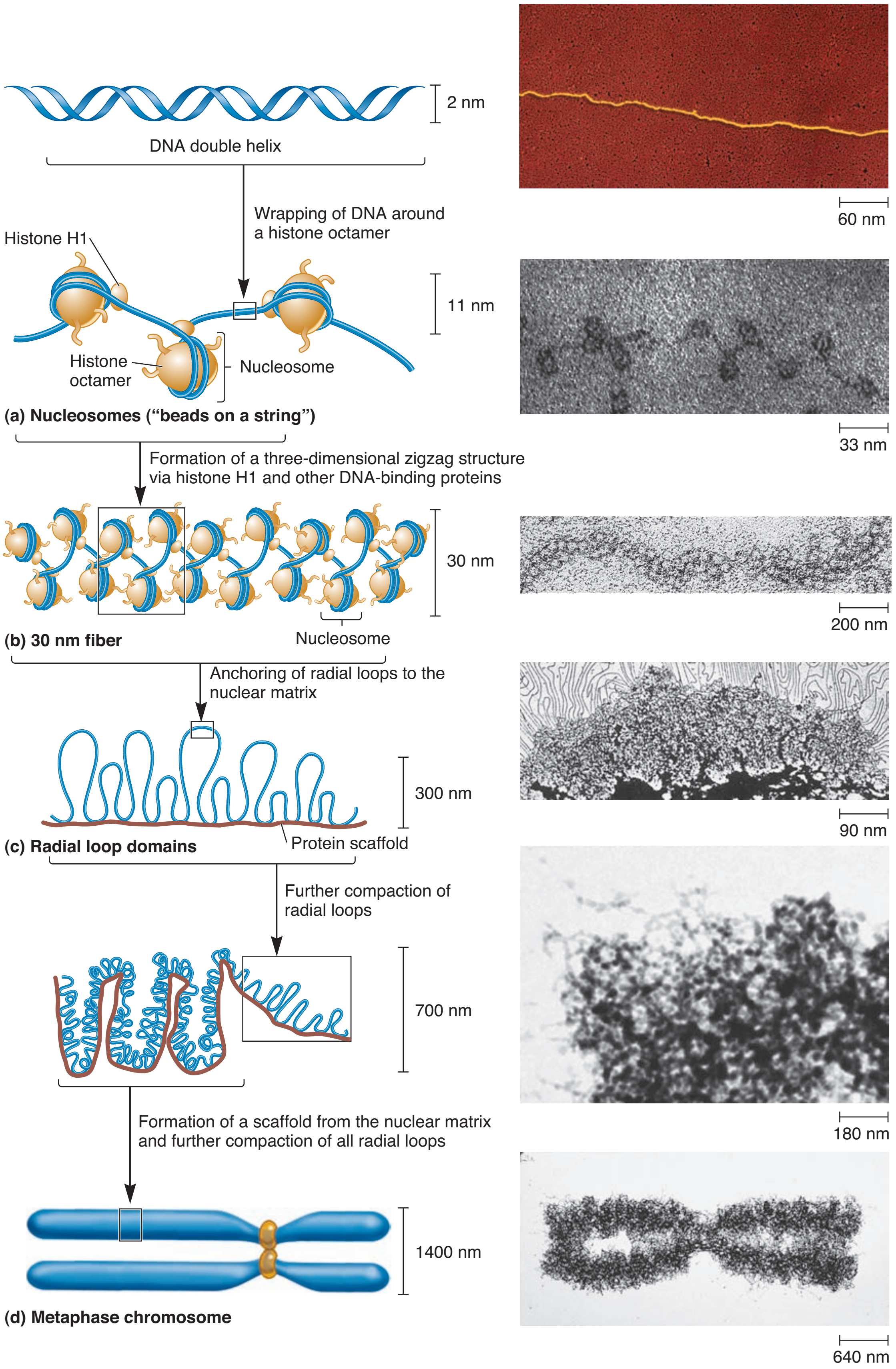

2 nm
DNA double helix
60 nm
Wrapping of DNA around a histone octamer
Histone H1
11 nm
Histone octamer
Nucleosome
(a) Nucleosomes ("beads on a string")
33 nm
Formation of a three-dimensional zigzag structure via histone H1 and other DNA-binding proteins
30 nm
200 nm
(b) 30 nm fiber
Nucleosome
Anchoring of radial loops to the nuclear matrix
300 nm
90 nm
(c) Radial loop domains
Protein scaffold
Further compaction of radial loops
700 nm
180 nm
Formation of a scaffold from the nuclear matrix and further compaction of all radial loops
1400 nm
(d) Metaphase chromosome
640 nm

FIGURE 12.17 **The steps in eukaryotic chromosomal compaction leading to the metaphase chromosome.**

Concept check: *Describe what structural changes convert a region that is 300 nm diameter to one that is 700 nm in diameter.*

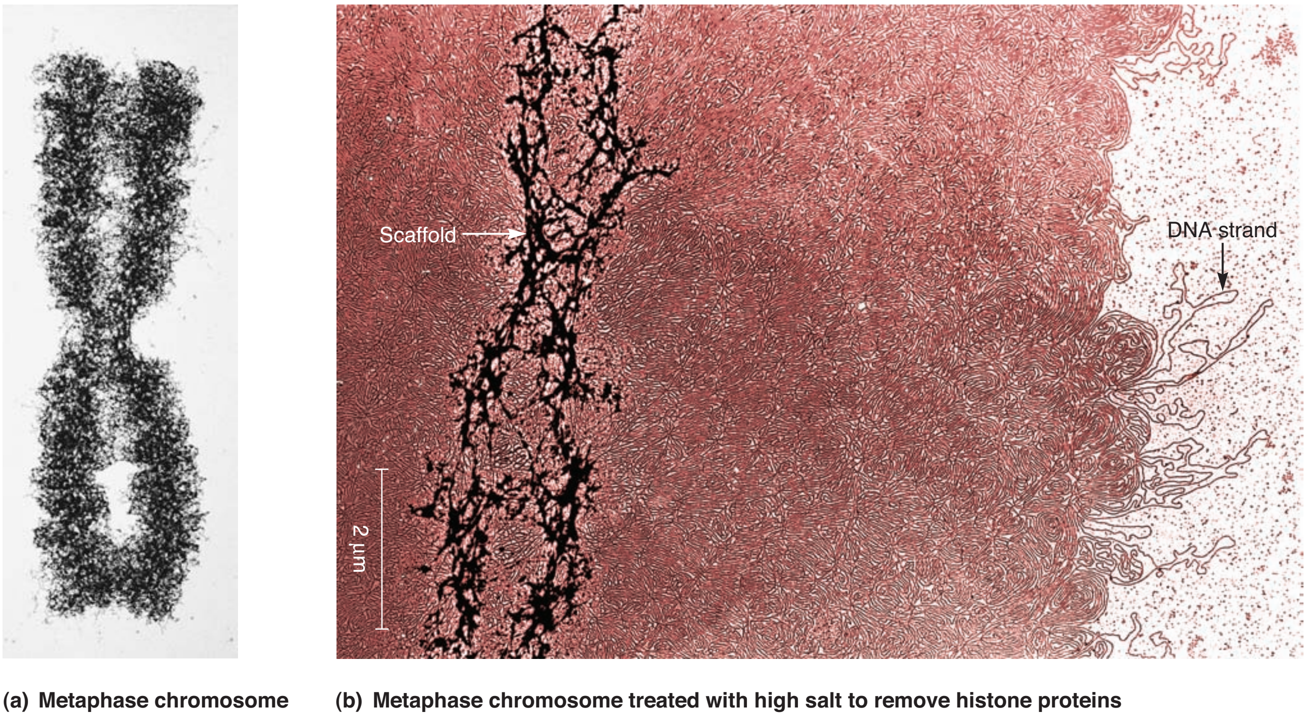

Scaffold
DNA strand
2 µm
(a) Metaphase chromosome
(b) Metaphase chromosome treated with high salt to remove histone proteins

FIGURE 12.18 **The importance of histone proteins and scaffolding proteins in the compaction of eukaryotic chromosomes.** **(a)** A metaphase chromosome. **(b)** A metaphase chromosome following treatment with high salt concentration to remove the histone proteins. The arrow on the left points to the scaffold (composed of nonhistone proteins), which anchors the bases of the radial loops. The arrow on the right points to an elongated strand of DNA.

Condensin and Cohesin Promote the Formation of Metaphase Chromosomes

Researchers are trying to understand the steps that lead to the formation and organization of metaphase chromosomes. During the past several years, studies in yeast and frog oocytes have been aimed at the identification of proteins that promote the conversion of interphase chromosomes into metaphase chromosomes. In yeast, mutants have been characterized that have alterations in the condensation or the segregation of chromosomes. Similarly, biochemical studies using frog oocytes resulted in the purification of protein complexes that promote chromosomal compaction or sister chromatid alignment. These two lines of independent research produced the same results. Researchers found that cells contain two multiprotein complexes called **condensin** and **cohesin,** which play critical roles in chromosomal compaction and sister chromatid alignment, respectively.

Condensin and cohesin are two completely distinct complexes, but both contain a category of proteins called **SMC proteins.** SMC stands for structural maintenance of chromosomes. These proteins use energy from ATP to catalyze changes in chromosome structure. Together with topoisomerases, SMC proteins have been shown to promote major changes in DNA structure. An emerging theme is that SMC proteins actively fold, tether, and manipulate DNA strands. They are dimers that have a V-shaped structure. The monomers, which are connected at a hinge region, have two long coiled arms with a head region that binds ATP (Figure 12.19). The length of each monomer is about 50 nm, which is equivalent to approximately 150 bp of DNA.

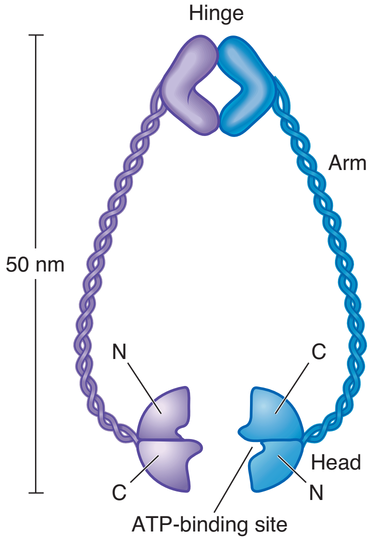


FIGURE 12.19 **The structure of SMC proteins.** This figure shows the generalized structure of SMC proteins, which are dimers consisting of hinge, arm, and head regions. The head regions bind and hydrolyze ATP. Condensin and cohesin have additional protein subunits not shown here. N indicates the amino terminus. C indicates the carboxyl terminus.

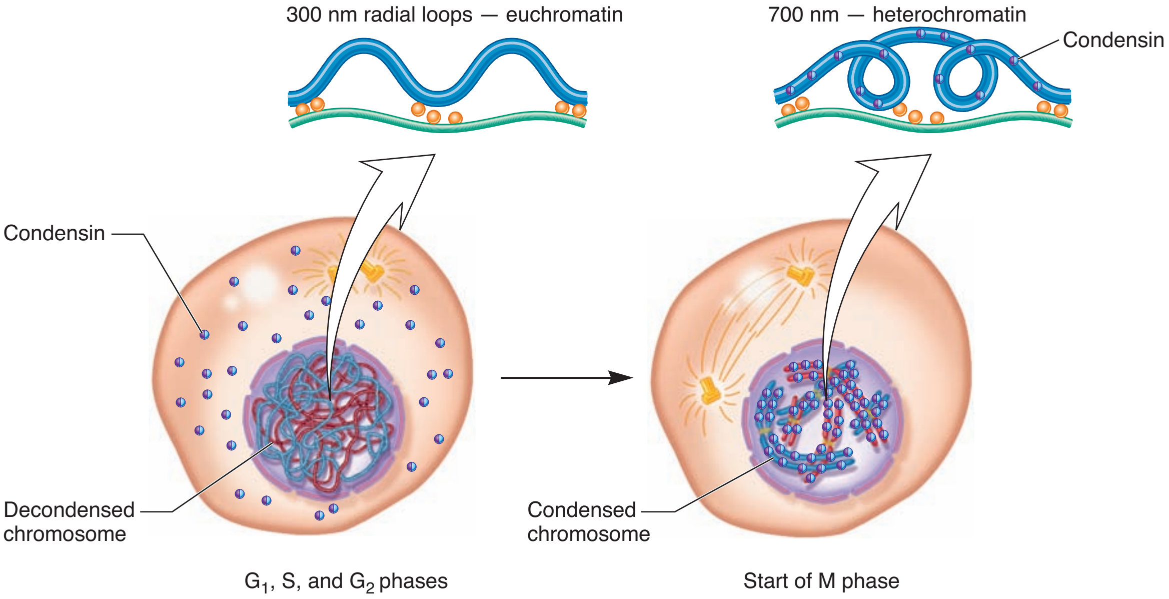


FIGURE 12.20 **The localization of condensin during interphase and the start of M phase.** During interphase (G_1, S, and G_2), most of the condensin protein is found outside the nucleus. The interphase chromosomes are largely euchromatic. At the start of M phase, condensin travels into the nucleus and binds to the chromosomes, which become heterochromatic due to a greater compaction of the radial loop domains.

As their names suggest, condensin and cohesin play different roles in metaphase chromosome structure. Prior to M phase, condensin is found outside the nucleus (**Figure 12.20**). However, as M phase begins, condensin is observed to coat the individual chromatids as euchromatin is converted into heterochromatin. The role of condensin in the compaction process is not well understood. Although condensin is often implicated in the process of chromosomal condensation, researchers have been able to deplete condensin from actively dividing cells, and the chromosomes are still able to condense. However, such condensed chromosomes show abnormalities in their ability to separate from each other during cell division. These results suggest that condensin is important in the proper organization of highly condensed chromosomes, such as those found during metaphase.

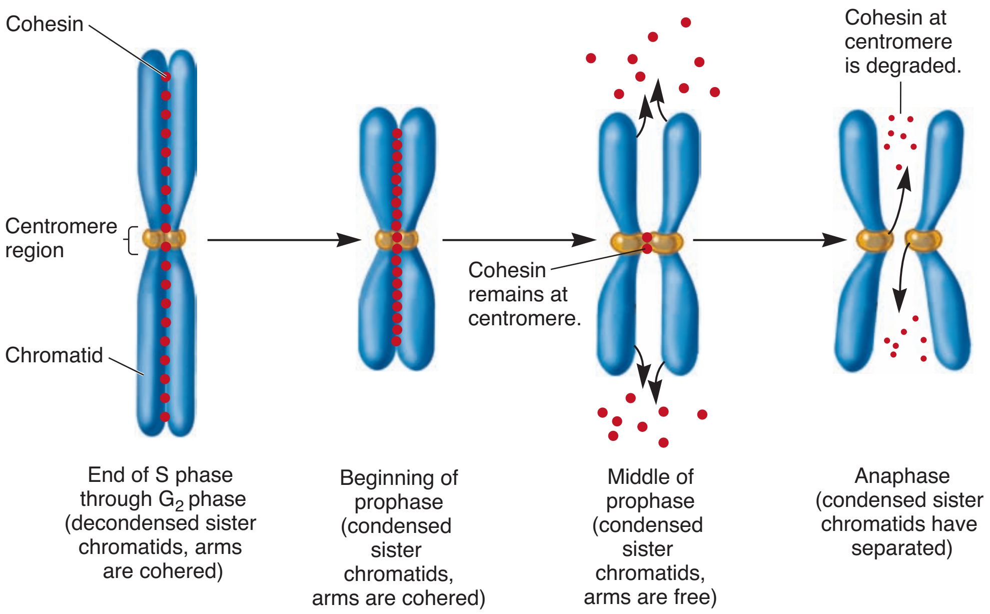
In comparison, the function of cohesin is to promote the binding (i.e., cohesion) between sister chromatids. After S phase and until the middle of prophase, sister chromatids remain attached to each other along their length. As shown in **Figure 12.21**, this attachment is promoted by cohesin, which is found along the entire length of each chromatid. In certain species, such as mammals, cohesins located along the chromosome arms are released during prophase, which allows the arms to separate. However, some cohesins remain attached, primarily to the centromeric regions, leaving the centromeric region as the main linkage before anaphase. At the start of anaphase, the cohesins bound to the centromere are rapidly degraded by a protease aptly named separase, thereby allowing sister chromatid separation.

12.6 REVIEWING THE KEY CONCEPTS

- Chromatin compaction occurs due to nucleosome formation and radial loop formation. A metaphase chromosome is completely heterochromatic due to the further compaction of radial loops (see Figure 12.17).
- Both histone and nonhistone proteins are important for the compaction of metaphase chromosomes (see Figure 12.18).
- Condensin and cohesin are SMC proteins that promote chromosome compaction and sister chromatid cohesion, respectively (see Figures 12.19–12.21).

12.6 COMPREHENSION QUESTIONS

1. The compaction leading to a metaphase chromosome involves which of the following?
   a. The formation of nucleosomes
   b. The formation of the 30-nm fiber
   c. Anchoring and further compaction of the radial loops
   d. All of the above
2. The role of cohesin is to
   a. make chromosomes more compact.
   b. allow for the replication of chromosomes.
   c. hold sister chromatids together.
   d. promote the separation of sister chromatids.


Cohesin
Centromere region
Chromatid
End of S phase through G2 phase (decondensed sister chromatids, arms are cohered)
Beginning of prophase (condensed sister chromatids, arms are cohered)
Cohesin remains at centromere.
Middle of prophase (condensed sister chromatids, arms are free)
Cohesin at centromere is degraded.
Anaphase (condensed sister chromatids have separated)

ONLINE ANIMATION

FIGURE 12.21 **The alignment of sister chromatids via cohesin.** After S phase is completed, many cohesin complexes bind along each chromatid, thereby facilitating their attachment to each other. During the middle of prophase, cohesin is released from the chromosome arms, but some cohesin remains in the centromeric regions. At the beginning of anaphase, the remaining cohesin complexes are rapidly degraded by separase, which promotes sister chromatid separation.

Concept check: Describe what happens to cohesin from the beginning of prophase to anaphase.

KEY TERMS

Page 246. chromosomes, genome, structural genes, intergenic regions
Page 247. origin of replication, repetitive sequences, nucleoid
Page 248. loop domains, DNA supercoiling, topoisomers, DNA gyrase
Page 249. topoisomerase I
Page 251. centromeres, kinetochore, telomeres
Page 252. introns, sequence complexity, moderately repetitive sequences
Page 253. transposable elements (TEs), highly repetitive sequences, retroelement, tandem array, nucleus
Page 254. chromatin, nucleosome, histone proteins
Page 257. 30-nm fiber
Page 258. nuclear matrix, nuclear lamina, internal nuclear matrix, radial loop domains, matrix-attachment regions (MARs), chromosome territory
Page 259. heterochromatin, euchromatin, constitutive heterochromatin, facultative heterochromatin
Page 260. scaffold
Page 262. condensin, cohesin, SMC proteins

CHAPTER SUMMARY

- Chromosomes contain the genetic material, which is DNA. The complete genetic complement of a cell or species is called a genome.

12.1 Organization of Sites Along Bacterial Chromosomes

- Bacterial chromosomes are typically circular and carry an origin of replication and a few thousand genes (see Figure 12.1).

12.2 Structure of Bacterial Chromosomes

- Bacterial chromosomes are found in the nucleoid region of a bacterial cell (see Figure 12.2).
- Bacterial chromosomes are made more compact by the formation of loop domains and DNA supercoiling (see Figures 12.3, 12.4).
- Negative DNA supercoiling can promote DNA strand separation (see Figure 12.5).

- DNA gyrase (topoisomerase II) is a bacterial enzyme that introduces negative supercoils (see Figure 12.6).
- Topoisomerase I relaxes negative supercoils.

12.3 Organization of Sites Along Eukaryotic Chromosomes

- Eukaryotic chromosomes are usually linear and contain multiple origins of replication, a centromere, telomeres, repetitive sequences, and many genes (see Figure 12.7).

12.4 Sizes of Eukaryotic Genomes and Repetitive Sequences

- The genome sizes of eukaryotes vary greatly. Some of this variation is due to the accumulation of repetitive sequences (see Figure 12.8).
- The human genome contains about 41% unique DNA sequences and 59% repetitive DNA sequences (see Figure 12.9).

12.5 Structure of Eukaryotic Chromosomes in Nondividing Cells

- The term *chromatin* refers to the DNA–protein complex found within eukaryotic chromosomes.
- Eukaryotic DNA wraps around an octamer of histone proteins to form nucleosomes (see Figure 12.10).
- Noll tested Kornberg's nucleosome model by digesting eukaryotic chromatin with varying concentrations of DNase I (see Figure 12.11).
- The linker histone, H1, plays a role in nucleosome compaction (see Figure 12.12).
- Nucleosomes are further compacted to form a 30-nm fiber. Solenoid and zigzag models have been proposed (see Figure 12.13).
- Chromatin is further compacted by the attachment of 30-nm fibers to protein filaments to form radial loop domains (see Figure 12.14).
- Within the cell nucleus, each eukaryotic chromosome occupies its own unique chromosome territory (see Figure 12.15).
- In nondividing cells, each chromosome has highly compacted regions called heterochromatin and less compacted regions called euchromatin (see Figure 12.16).

12.6 Structure of Eukaryotic Chromosomes During Cell Division

- Chromatin compaction occurs due to nucleosome formation and radial loop formation. A metaphase chromosome is completely heterochromatic due to the further compaction of radial loops (see Figure 12.17).
- Both histone and nonhistone proteins are important for the compaction of metaphase chromosomes (see Figure 12.18).
- Condensin and cohesin are SMC proteins that promote chromosome compaction and sister chromatid cohesion, respectively (see Figures 12.19–12.21).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Let's suppose a bacterial DNA molecule is given a left-handed twist. How does this affect the structure and function of the DNA?

Answer: A left-handed twist is negative supercoiling. Negative supercoiling makes the bacterial chromosome more compact. It also promotes DNA functions that involve strand separation, including gene transcription and DNA replication.

S2. Describe the characteristics of highly repetitive DNA.

Answer: Highly repetitive DNA, as its name suggests, is a DNA sequence that is repeated many times, from tens of thousands to millions of times throughout the genome. It can be interspersed in the genome or found clustered in a tandem array, in which a short nucleotide sequence is repeated many times in a row.

S3. To hold bacterial DNA in a more compact configuration, specific proteins must bind to the DNA and stabilize its conformation (as shown in Figure 12.3). Several different proteins are involved in this process. These proteins have been collectively referred to as "histone-like" due to their functional similarity to the histone proteins found in eukaryotes. Based on your knowledge of eukaryotic histone proteins, what biochemical properties would you expect from bacterial histone-like proteins?

Answer: The histone-like proteins have the properties expected for proteins involved in DNA folding. They are all small proteins found in relative abundance within the bacterial cell. In some cases, the histone-like proteins are biochemically similar to eukaryotic histones. For example, they tend to be basic (positively charged) and bind to DNA in a non-sequence-dependent fashion. In addition, other proteins appear to bind to bacterial DNA at specific sites in order to promote DNA bending.

S4. If you assume the average length of linker DNA is 50 bp, approximately how many nucleosomes are found in the haploid human genome, which contains 3 billion bp?

Answer: Assuming a size of 3 billion bp, and that 146 bp wrap around a histone octamer, with 50 bp in the intervening region:

3,000,000,000/196 = 15,306,122, or about 15.3 million

Conceptual Questions

C1. What is a bacterial nucleoid? With regard to cellular membranes, what is the difference between a bacterial nucleoid and a eukaryotic nucleus?

C2. In Part II of this textbook, we considered inheritance patterns for diploid eukaryotic species. Bacteria frequently contain two or more nucleoids. With regard to genes and alleles, how is a bacterium that contains two nucleoids similar to a diploid eukaryotic cell, and how is it different?

C3. Describe the two main mechanisms by which the bacterial DNA becomes compacted.

C4. As described in Chapter 11, 1 bp of DNA is approximately 0.34 nm in length. A bacterial chromosome is about 4 million bp in length and is organized into about 100 loops that are about 40,000 bp in length.

A. If it was stretched out linearly, how long (in micrometers) would one loop be?

B. If a bacterial chromosomal loop is circular, what would be its diameter? (*Note:* Circumference = πD, where D is the diameter of the circle.)

C. Is the diameter of the circular loop calculated in part B small enough to fit inside a bacterium? The dimensions of the bacterial cytoplasm, such as *E. coli,* are roughly 0.5 μm wide and 1.0 μm long.

C5. Why is DNA supercoiling called supercoiling rather than just coiling? Why is positive supercoiling called overwinding and negative supercoiling called underwinding? How would you define the terms *positive* and *negative supercoiling* for Z DNA (described in Chapter 11)?

C6. Coumarins and quinolones are two classes of drugs that inhibit bacterial growth by directly inhibiting DNA gyrase. Discuss two reasons why inhibiting DNA gyrase might inhibit bacterial growth.

C7. Take two pieces of string that are approximately 10 inches each, and create a double helix by wrapping them around each other to make 10 complete turns. Tape one end of the strings to a table, and now twist the strings three times (360° each time) in a right-handed direction. *Note:* As you are looking down at the strings from above, a right-handed twist is in the clockwise direction.

A. Did the three turns create more or fewer turns in your double helix? How many turns are now in your double helix?

B. Is your double helix right-handed or left-handed? Explain your answer.

C. Did the three turns create any supercoils?

D. If you had coated your double helix with rubber cement and allowed the cement to dry before making the three additional right-handed turns, would the rubber cement make it more or less likely for the three turns to create supercoiling? Would a pair of cemented strings be more or less like a real DNA double helix than an uncemented pair of strings? Explain your answer.

C8. Try to explain the function of DNA gyrase with a drawing.

C9. How are two topoisomers different from each other? How are they the same?

C10. On rare occasions, a chromosome can suffer a small deletion that removes the centromere. When this occurs, the chromosome usually is not found within subsequent daughter cells. Explain why a chromosome without a centromere is not transmitted very efficiently from mother to daughter cells. (*Note:* If a chromosome is located outside the nucleus after telophase, it is degraded.)

C11. What is the function of a centromere? At what stage of the cell cycle would you expect the centromere to be the most important?

C12. Describe the structures of a nucleosome and a 30-nm fiber.

C13. Beginning with the G_1 phase of the cell cycle, describe the level of compaction of the eukaryotic chromosome. How does the level of compaction change as the cell progresses through the cell cycle? Why is it necessary to further compact the chromatin during mitosis?

C14. Draw a picture depicting the binding between the nuclear matrix and MARs.

C15. Compare heterochromatin and euchromatin. What are the differences between them?

C16. Compare the structure and cell localization of chromosomes during interphase and M phase.

C17. What types of genetic activities occur during interphase? Explain why these activities cannot occur during M phase.

C18. Let's assume the linker DNA averages 54 bp in length. How many molecules of H2A would you expect to find in a DNA sample that is 46,000 bp in length?

C19. In Figure 12.12, what are we looking at in part b? Is this an 11-nm fiber, a 30-nm fiber, or a 300-nm fiber? Does this DNA come from a cell during M phase or interphase?

C20. What are the roles of the core histone proteins compared with the role of histone H1 in the compaction of eukaryotic DNA?

C21. A typical eukaryotic chromosome found in humans contains about 100 million bp of DNA. As described in Chapter 11, 1 bp of DNA has a linear length of 0.34 nm.

A. What is the linear length of the DNA for a typical human chromosome in micrometers?

B. What is the linear length of a 30-nm fiber of a typical human chromosome?

C. Based on your calculation of part B, would a typical human chromosome fit inside the nucleus (with a diameter of 5 μm) if the 30-nm fiber were stretched out in a linear manner? If not, explain how a typical human chromosome fits inside the nucleus during interphase.

C22. Which of the following terms should not be used to describe a Barr body?

A. Chromatin

B. Euchromatin

C. Heterochromatin

D. Chromosome

E. Genome

C23. Discuss the differences in the compaction levels of metaphase chromosomes compared with interphase chromosomes. When would you expect gene transcription and DNA replication to take place, during M phase or interphase? Explain why.

C24. What is an SMC protein? Describe two examples.

Application and Experimental Questions

E1. Two circular DNA molecules, which we will call molecule A and molecule B, are topoisomers of each other. When viewed under the electron microscope, molecule A appears more compact than molecule B. The level of gene transcription is much lower for molecule A. Which of the following three possibilities could account for these observations?

First possibility: Molecule A has three positive supercoils, and molecule B has three negative supercoils.

Second possibility: Molecule A has four positive supercoils, and molecule B has one negative supercoil.

Third possibility: Molecule A has zero supercoils, and molecule B has three negative supercoils.

E2. Let's suppose that you have isolated DNA from a cell and have viewed it under a microscope. It looks supercoiled. What experiment would you perform to determine if it is positively or negatively supercoiled? In your answer, describe your expected results. You may assume that you have purified topoisomerases at your disposal.

E3. We seem to know more about the structure of eukaryotic chromosomal DNA than bacterial DNA. Discuss why you think this is so, and list several experimental procedures that have yielded important information concerning the compaction of eukaryotic chromatin.

E4. When chromatin is treated with a moderate salt concentration, the linker histone H1 is removed (see Figure 12.12a). Higher salt concentration removes the rest of the histone proteins (see Figure 12.18b). If the experiment of Figure 12.11 were carried out after the DNA was treated with moderate or high salt, what would be the expected results?

E5. Let's suppose you have isolated chromatin from some bizarre eukaryote with a linker region that is usually 300 to 350 bp in length. The nucleosome structure is the same as in other eukaryotes. If you digested this eukaryotic organism's chromatin with a high concentration of DNase I, what would be your expected results?

E6. If you were given a sample of chromosomal DNA and asked to determine if it is bacterial or eukaryotic, what experiment would you perform, and what would be your expected results?

E7. Consider how histone proteins bind to DNA and then explain why a high salt concentration can remove histones from DNA (as shown in Figure 12.18b).

E8. In Chapter 21, the technique of fluorescence in situ hybridization (FISH) is described. This is another method for examining sequence complexity within a genome. In this technique, a particular DNA sequence, such as a particular gene sequence, can be detected within an intact chromosome by using a DNA probe that is complementary to the sequence. For example, let's consider the β-globin gene, which is found on human chromosome 11. A probe complementary to the β-globin gene binds to the β-globin gene and shows up as a brightly colored spot on human chromosome 11. In this way, researchers can detect where the β-globin gene is located within a set of chromosomes. Because the β-globin gene is unique and because human cells are diploid (i.e., have two copies of each chromosome), a FISH experiment shows two bright spots per cell; the probe binds to each copy of chromosome 11. What would you expect to see if you used the following types of probes?

A. A probe complementary to the *Alu*I sequence

B. A probe complementary to a tandemly repeated sequence near the centromere of the X chromosome

Questions for Student Discussion/Collaboration

1. Bacterial and eukaryotic chromosomes are very compact. Discuss the advantages and disadvantages of having a compact structure.
2. The prevalence of highly repetitive sequences seems rather strange to many geneticists. Do they seem strange to you? Why or why not? Discuss whether or not you think they have an important function.
3. Discuss and make a list of the similarities and differences between bacterial and eukaryotic chromosomes.

Answers to Comprehension Questions

12.1: d

12.2: d, b, d

12.3: e

12.4: a

12.5: b, b, d, c

12.6: d, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

13

CHAPTER OUTLINE

13.1 Structural Overview of DNA Replication

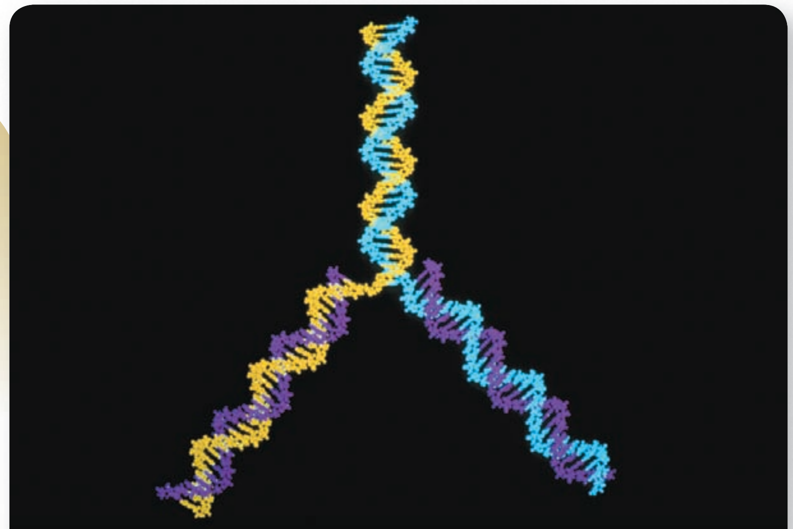
13.2 Bacterial DNA Replication: The Formation of Two Replication Forks at the Origin of Replication

13.3 Bacterial DNA Replication: Synthesis of New DNA Strands

13.4 Bacterial DNA Replication: Chemistry and Accuracy

13.5 Eukaryotic DNA Replication

13.6 Homologous Recombination



A model for DNA undergoing replication. *This molecular model shows a DNA replication fork, the site where new DNA strands are made. In this model, the original DNA is yellow and blue. The newly made strands are purple.*

DNA REPLICATION AND RECOMBINATION

As discussed throughout Chapters 2 to 10, genetic material is transmitted from parent to offspring and from cell to cell. For transmission to occur, the genetic material must be copied. During this process, known as **DNA replication,** the original DNA strands are used as templates for the synthesis of new DNA strands. We will begin Chapter 13 with a consideration of the structural features of the double helix that underlie the replication process. Next we examine how chromosomes are replicated within living cells, addressing the following questions: where does DNA replication begin, how does it proceed, and where does it end? We first consider how DNA replication occurs within bacterial cells, and then turn our attention to the unique features of the replication of eukaryotic DNA. At the molecular level, it is rather remarkable that the replication of chromosomal DNA occurs very quickly, very accurately, and at the appropriate time in the life of the cell. For this to happen, many cellular proteins play vital roles. In this chapter, we will examine the mechanism of DNA replication and consider the functions of several proteins involved in the process.

We will also examine the molecular mechanism by which segments of DNA, such as those found in chromosomes, become rearranged. During **homologous recombination,** DNA segments, which are similar or identical to each other, break and rejoin to form a new combination. As described in Chapter 2, homologous recombination occurs when chromosomes cross over during meiosis. Not only does homologous recombination enhance genetic diversity, it also helps to repair DNA and ensures the proper segregation of chromosomes.

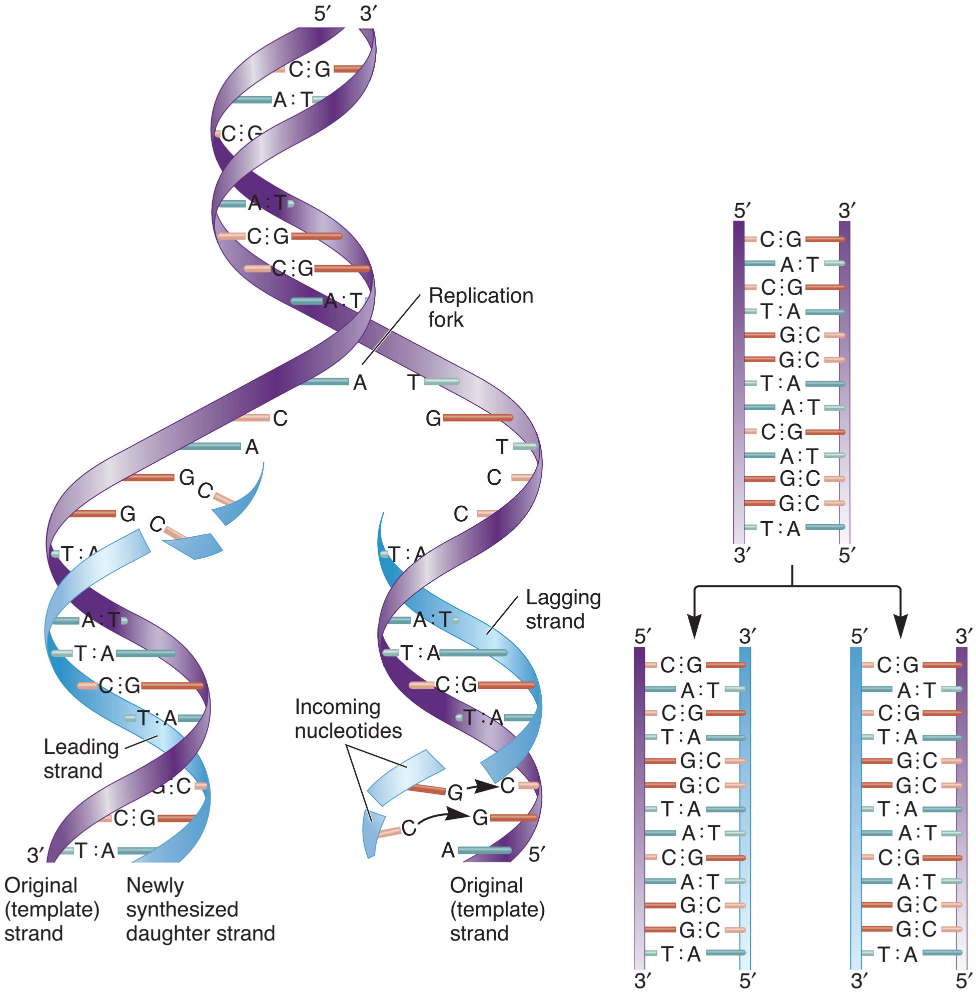
13.1 STRUCTURAL OVERVIEW OF DNA REPLICATION

Because they bear directly on the replication process, let's begin by recalling a few important structural features of the double helix from Chapter 11. The double helix is composed of two DNA strands. The individual building blocks of each strand are nucleotides, which contain one of four bases: adenine, thymine, guanine, or cytosine. The double-stranded structure is held together by base stacking and by hydrogen bonding between the bases in opposite strands. A critical feature of the double-helix structure is that adenine (A) hydrogen bonds with thymine (T), and guanine (G) hydrogen bonds with cytosine (C). This rule, known as the AT/GC rule, is the basis for the complementarity of the base sequences in double-stranded DNA.

Another feature worth noting is that the strands within a double helix have an antiparallel alignment. This directionality is determined by the orientation of sugar molecules within the sugar–phosphate backbone. If one strand is running in the 5′ to 3′ direction, the complementary strand is running in the 3′ to 5′ direction. The issue of directionality becomes important when we consider the function of the enzymes that synthesize new DNA strands. In this section, we will consider how the structure of the DNA double helix provides the basis for DNA replication.

Existing DNA Strands Act as Templates for the Synthesis of New Strands

As shown in Figure 13.1a, DNA replication relies on the complementarity of DNA strands according to the AT/GC rule. During the replication process, the two complementary strands of DNA come apart and serve as **template strands,** or **parental strands,** for the synthesis of two new strands of DNA. After the double helix has separated, individual nucleotides have access to the template strands. Hydrogen bonding between individual nucleotides and the template strands must obey the AT/GC rule.



(a) The mechanism of DNA replication

(b) The products of replication

ONLINE ANIMATION

FIGURE 13.1 **The structural basis for DNA replication.** **(a)** The mechanism of DNA replication as originally proposed by Watson and Crick. As we will see, the synthesis of one newly made strand (the leading strand) occurs in the direction toward the replication fork, whereas the synthesis of the other newly made strand (the lagging strand) occurs in small segments away from the replication fork. **(b)** DNA replication produces two copies of DNA with the same sequence as the original DNA molecule.

Concept check: *What features of DNA structure enable it to be replicated?*

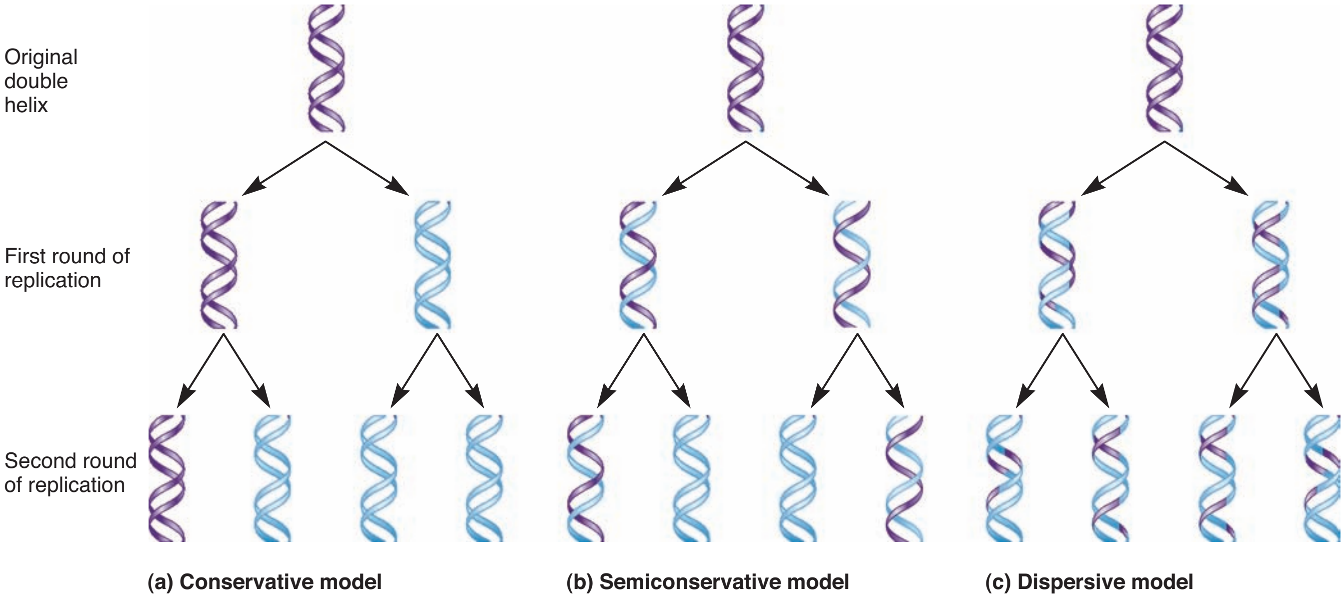

Original double helix
First round of replication
Second round of replication
(a) Conservative model
(b) Semiconservative model
(c) Dispersive model

FIGURE 13.2 **Three possible models for DNA replication.** The two original parental DNA strands are shown in purple, and the newly made strands after one and two generations are shown in light blue.

Concept check: *Which of these models is the correct one?*

To complete the replication process, a covalent bond is formed between the phosphate of one nucleotide and the sugar of the previous nucleotide. The two newly made strands are referred to as the **daughter strands.** Note that the base sequences are identical in both double-stranded molecules after replication (**Figure 13.1b**). Therefore, DNA is replicated so both copies retain the same information—the same base sequence—as the original molecule.

Three Different Models Were Proposed That Described the Net Result of DNA Replication

Scientists in the late 1950s considered three different mechanisms to explain the net result of DNA replication. These mechanisms are shown in **Figure 13.2.** The first is referred to as a **conservative model.** According to this hypothesis, both strands of parental DNA remain together following DNA replication. In this model, the original arrangement of parental strands is completely conserved, while the two newly made daughter strands also remain together following replication. The second is called a **semiconservative model.** In this mechanism, the double-stranded DNA is half conserved following the replication process. In other words, the newly made double-stranded DNA contains one parental strand and one daughter strand. The third, called the **dispersive model,** proposes that segments of parental DNA and newly made DNA are interspersed in both strands following the replication process.

The semiconservative model, shown in Figure 13.2b, is the correct mechanism. How did scientists determine this? In 1958, Matthew Meselson and Franklin Stahl devised a method to experimentally distinguish newly made daughter strands from the original parental strands. Their technique involved labeling DNA with a heavy isotope of nitrogen (^{15}N). Nitrogen, which is found within the bases of DNA, occurs in both a heavy and light (^{14}N) form. Prior to their experiment, they grew *E. coli* cells in the presence of ^{15}N for many generations. This produced a population of cells in which all of the DNA was heavy-labeled. At the start of their experiment, shown in **Figure 13.3** (generation 0), they switched the bacteria to a medium that contained only ^{14}N and then collected samples of cells after various time points. Under the growth conditions they employed, 30 minutes is the time required for one doubling, or one generation time. Because the bacteria were doubling in a medium that contained only ^{14}N, all of the newly made DNA strands were labeled with light nitrogen, but the original strands remained in the heavy form.

Meselson and Stahl then analyzed the density of the DNA by centrifugation, using a cesium chloride (CsCl) gradient. (The procedure of gradient centrifugation is described in the Appendix.) If both DNA strands contained ^{14}N, the DNA would have a light density and sediment near the top of the tube. If one strand contained ^{14}N and the other strand contained ^{15}N, the DNA would be half-heavy and have an intermediate density. Finally, if both strands contained ^{15}N, the DNA would be heavy and move closer to the bottom of the centrifuge tube.

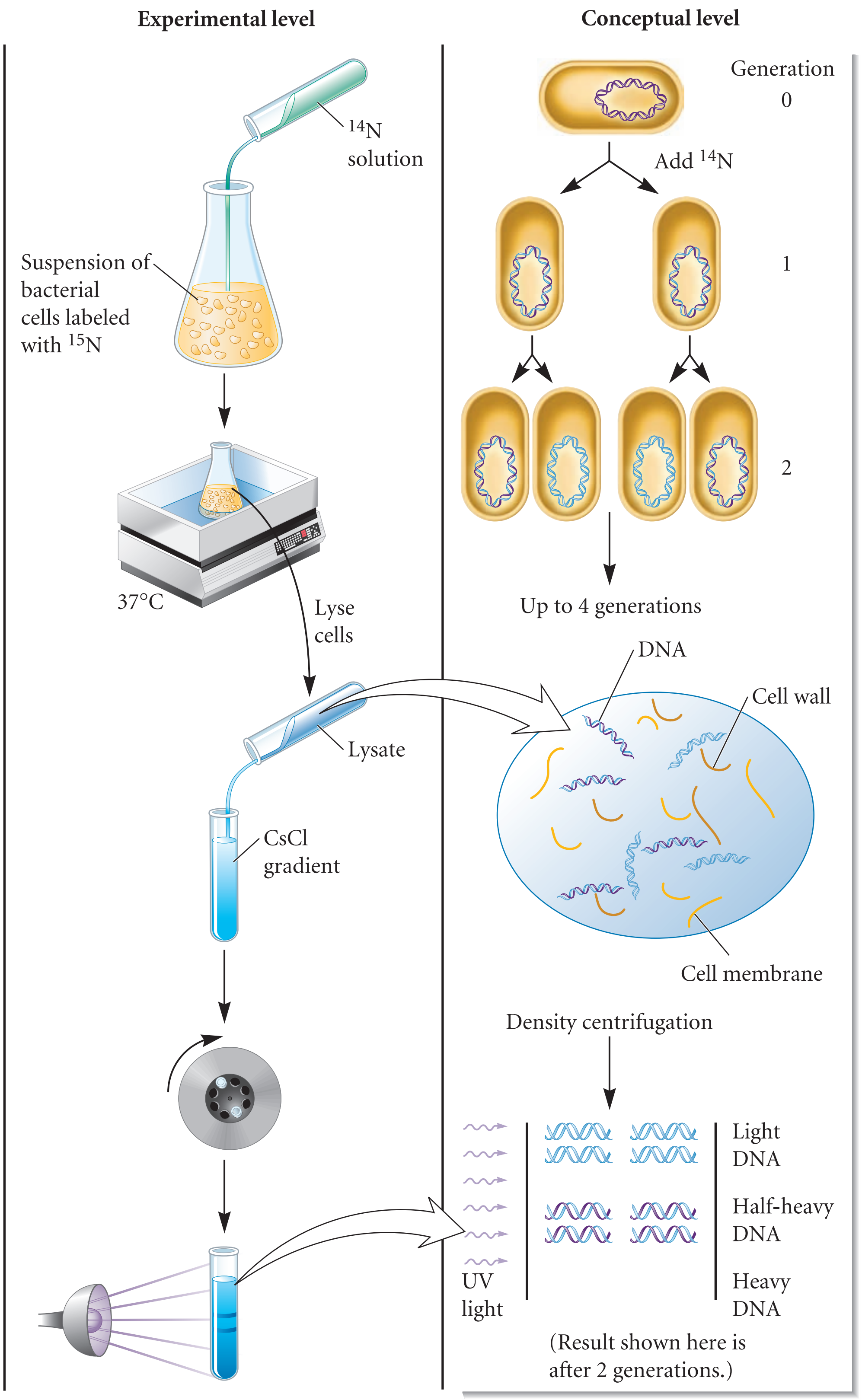
▶ THE HYPOTHESIS

Based on Watson's and Crick's ideas, the hypothesis was that DNA replication is semiconservative. Figure 13.2 also shows two alternative models.

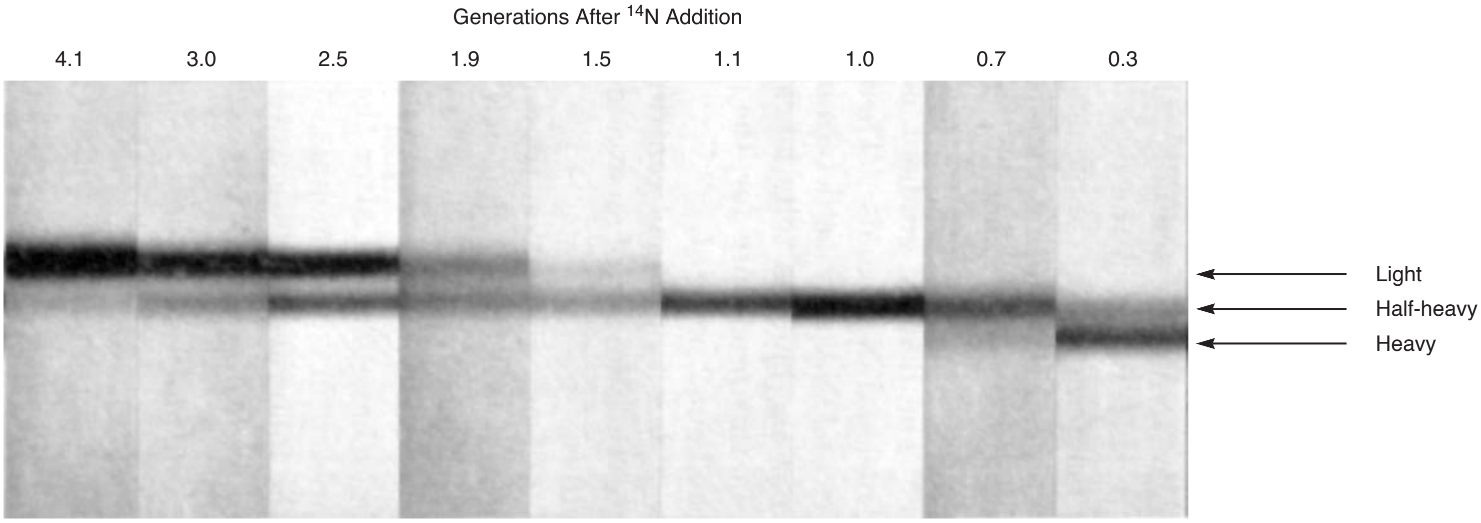
▶ **TESTING THE HYPOTHESIS — FIGURE 13.3** **Evidence that DNA replication is semiconservative.**

Starting material: A strain of *E. coli* that has been grown for many generations in the presence of ^{15}N. All of the nitrogen in the DNA is labeled with ^{15}N.

1. Add an excess of ^{14}N-containing compounds to the bacterial cells so all of the newly made DNA will contain ^{14}N.
2. Incubate the cells for various lengths of time. Note: The ^{15}N-labeled DNA is shown in purple and the ^{14}N-labeled DNA is shown in blue.
3. Lyse the cells by the addition of lysozyme and detergent, which disrupt the bacterial cell wall and cell membrane, respectively.
4. Load a sample of the lysate onto a CsCl gradient. (Note: The average density of DNA is around 1.7 g/cm^3, which is well isolated from other cellular macromolecules.)
5. Centrifuge the gradients until the DNA molecules reach their equilibrium densities.
6. DNA within the gradient can be observed under a UV light.


Experimental level
Conceptual level
^{14}N solution
Suspension of bacterial cells labeled with ^{15}N
37°C
Lyse cells
Lysate
CsCl gradient
Generation 0
Add ^{14}N
1
2
Up to 4 generations
DNA
Cell wall
Cell membrane
Density centrifugation
UV light
Light DNA
Half-heavy DNA
Heavy DNA
(Result shown here is after 2 generations.)

THE DATA



*Data from: Meselson, M. and Stahl, F.W. (1958) The Replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 44: 671–682

INTERPRETING THE DATA

As seen in the data of Figure 13.3, after one round of DNA replication (i.e., one generation), all of the DNA sedimented at a density that was half-heavy. Both the semiconservative and dispersive models are consistent with this result. In contrast, the conservative model predicts two separate DNA types: a light type and a heavy type. Because all of the DNA had sedimented as a single band, this model was disproved. According to the semiconservative model, the replicated DNA would contain one original strand (a heavy strand) and a newly made daughter strand (a light strand). Likewise, in a dispersive model, all of the DNA should have been half-heavy after one generation as well. To determine which of these two remaining models is correct, Meselson and Stahl had to investigate future generations.

After approximately two rounds of DNA replication (i.e., 1.9 generations), a mixture of light DNA and half-heavy DNA was observed. This result was consistent with the semiconservative model of DNA replication, because some DNA molecules should contain all light DNA, and other molecules should be half-heavy (see Figure 13.2b). The dispersive model predicts that after two generations, the heavy nitrogen would be evenly dispersed among four strands, each strand containing 1/4 heavy nitrogen and 3/4 light nitrogen (see Figure 13.2c). However, this result was not obtained. Instead, the results of the Meselson and Stahl experiment provided compelling evidence in favor of only the semiconservative model for DNA replication.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

13.1 REVIEWING THE KEY CONCEPTS

- DNA replication occurs when the strands of DNA unwind and each strand is used as a template to make a new strand according to the AT/GC rule. The resulting DNA molecules have the same base sequence as the original DNA (see Figure 13.1).
- By labeling DNA with heavy and light isotopes of nitrogen and using centrifugation, Meselson and Stahl showed that DNA replication is semiconservative (see Figures 13.2, 13.3).

13.1 COMPREHENSION QUESTIONS

1. The complementarity of DNA strands is based on the
   a. chemical properties of a phosphodiester linkage.
   b. the binding of proteins to the DNA.
   c. AT/GC rule.
   d. none of the above.
2. To make a new DNA strand, which of the following is necessary?
   a. A template strand
   b. Nucleotides
   c. Heavy nitrogen
   d. Both a and b
3. The model that correctly describes the process of DNA replication is the
   a. conservative model.
   b. semiconservative model.
   c. dispersive model.
   d. all of the above.

13.2 BACTERIAL DNA REPLICATION: THE FORMATION OF TWO REPLICATION FORKS AT THE ORIGIN OF REPLICATION

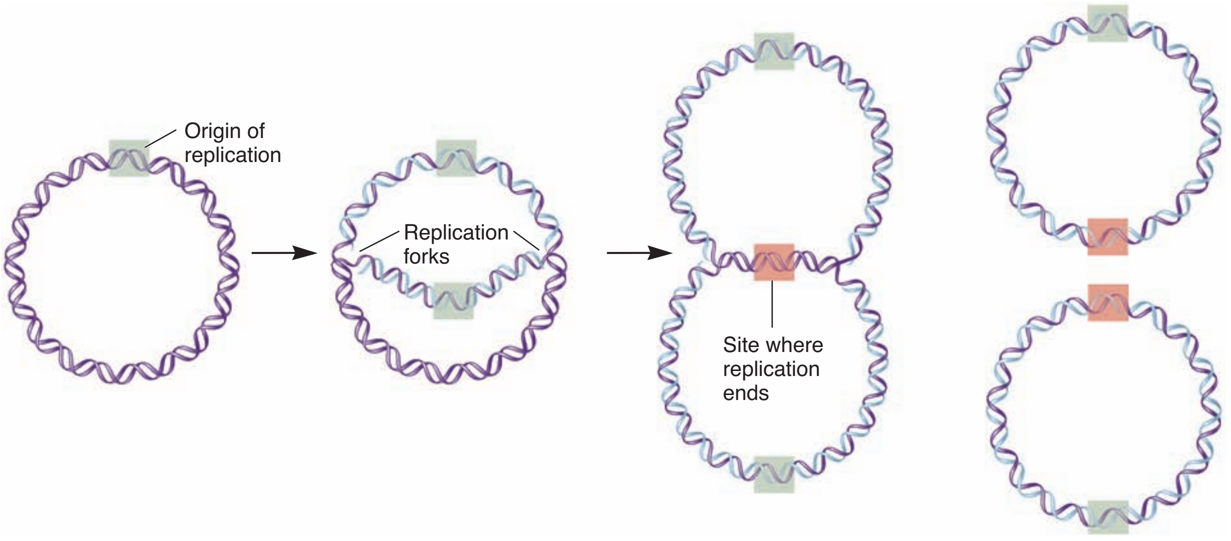
Thus far, we have considered how a complementary, double-stranded structure underlies the ability of DNA to be copied. In addition, the experiments of Meselson and Stahl showed that DNA replication results in two double helices, each one containing an original parental strand and a newly made daughter strand. We now turn our attention to how DNA replication actually occurs within living cells. Much research has focused on understanding the process in the bacterium *Escherichia coli.* In this section, we will examine how bacterial DNA replication begins.

Bacterial Chromosomes Contain a Single Origin of Replication

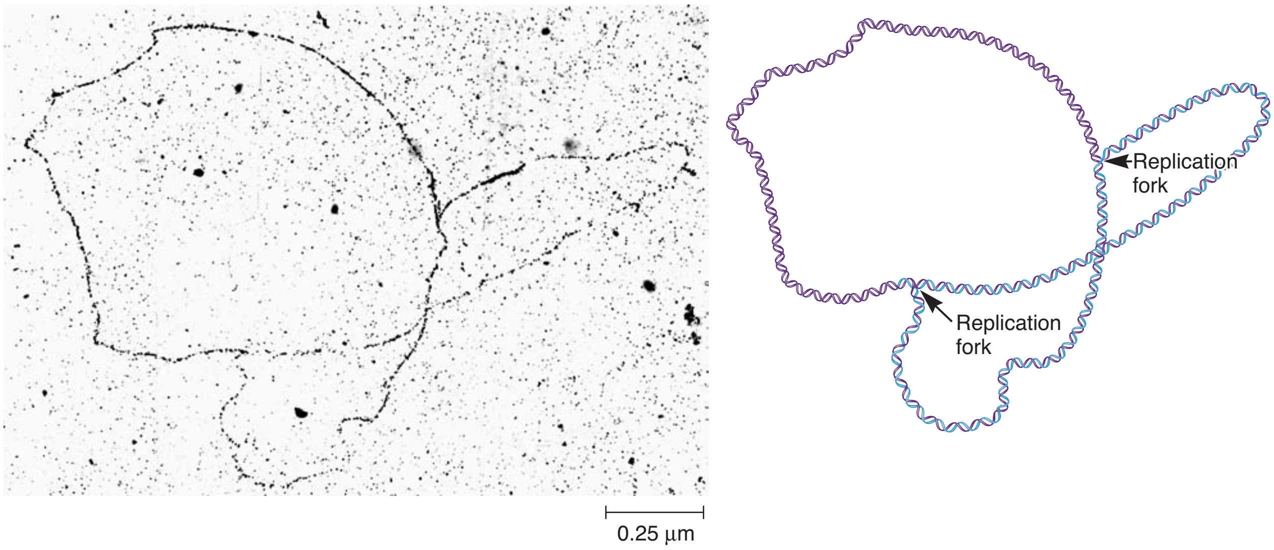
Figure 13.4 presents an overview of the process of bacterial chromosomal replication. The site on the bacterial chromosome where DNA synthesis begins is known as the **origin of replication.** Bacterial chromosomes have a single origin of replication. The synthesis of new daughter strands is initiated within the origin and proceeds in both directions, or **bidirectionally,** around the bacterial chromosome. A **replication fork** is the site where the parental DNA strands have separated and new daughter strands are being made. Two replication forks move in opposite directions outward from the origin and eventually meet each other on the opposite side of the bacterial chromosome when replication is completed.

Replication Is Initiated at the Origin of Replication

Considerable research has focused on the origin of replication in *E. coli.* This origin is named *oriC* for origin of Chromosomal



(a) Bacterial chromosome replication



(b) Autoradiograph of an *E. coli* chromosome in the act of replication

FIGURE 13.4 **The process of bacterial chromosome replication.** **(a)** An overview of the process of bacterial chromosome replication. **(b)** A replicating *E. coli* chromosome visualized by autoradiography and transmission electron microscopy (TEM). This chromosome was radiolabeled by growing bacterial cells in media containing radiolabeled thymidine. The diagram at the right shows the locations of the two replication forks. The chromosome is about one-third replicated. New strands are shown in blue.

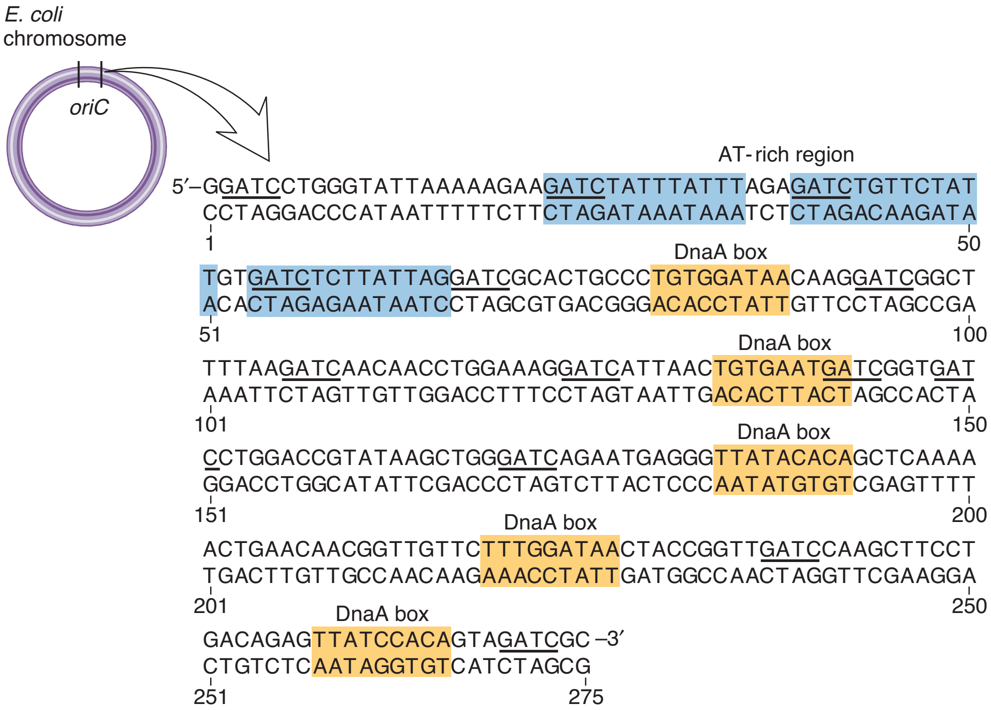

E. coli chromosome
oriC
AT-rich region
5′–GGATCCTGGGTATTAAAAAGAAGATCTATTTATTTAGAGATCTGTTCTAT
CCTAGGACCCATAATTTTTCTTCTAGATAAATAAATCTCTAGACAAGATA
1
50
DnaA box
TGTGATCTCTTATTAGGATCGCACTGCCCTGTGGATAACAAGGATCGGCT
ACACTAGAGAATAATCCTAGCGTGACGGGACACCTATTGTTCCTAGCCGA
51
100
DnaA box
TTTAAGATCAACAACCTGGAAAGGATCATTAACTGTGAATGATCGGTGAT
AAATTCTAGTTGTTGGACCTTTCCTAGTAATTGACACTTACTAGCCACTA
101
150
DnaA box
CCTGGACCGTATAAGCTGGGATCAGAATGAGGGTTATACACAGCTCAAAA
GGACCTGGCATATTCGACCCTAGTCTTACTCCCAATATGTGTCGAGTTTT
151
200
DnaA box
ACTGAACAACGGTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCT
TGACTTGTTGCCAACAAGAAACCTATTGATGGCCAACTAGGTTCGAAGGA
201
250
DnaA box
GACAGAGTTATCCACAGTAGATCGC–3′
CTGTCTCAATAGGTGTCATCTAGCG
251
275

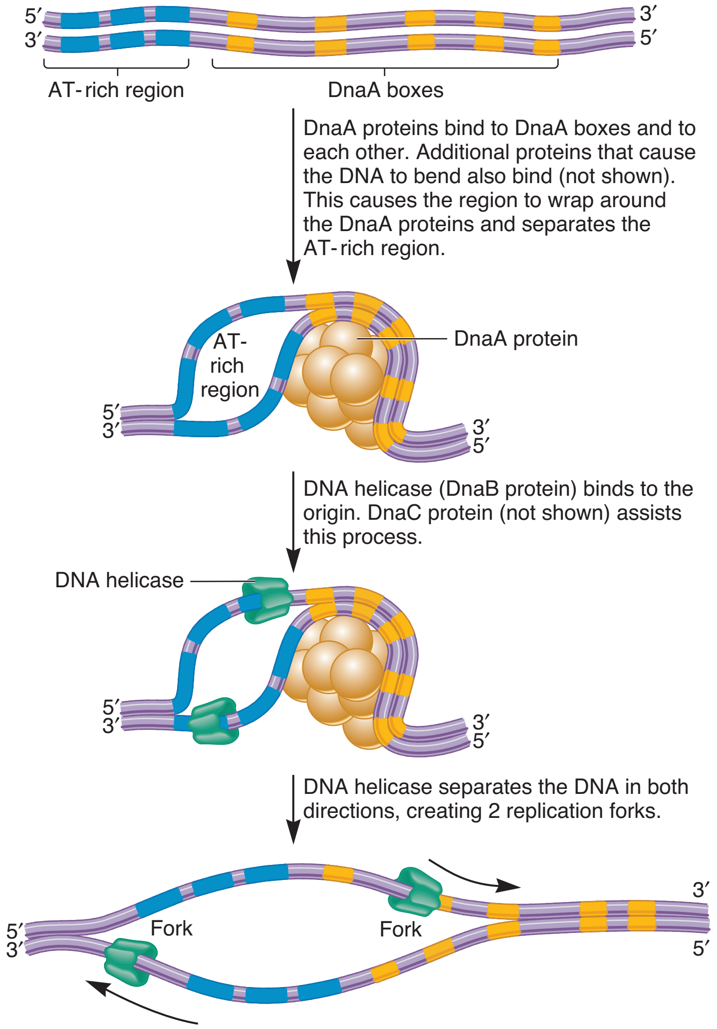
FIGURE 13.5 **The sequence of *oriC* in *E. coli.*** The AT-rich region is composed of three tandem repeats that are 13 bp long and highlighted in blue. The five DnaA boxes are highlighted in orange. The GATC methylation sites are underlined.

Concept check: What are the functions of the AT-rich region and DnaA boxes?

replication (**Figure 13.5**). Three types of DNA sequences are found within *oriC:* an AT-rich region, DnaA box sequences, and GATC methylation sites.

DNA replication is initiated by the binding of **DnaA proteins** to sequences within the origin known as **DnaA box sequences.** The DnaA box sequences serve as recognition sites for the binding of the DnaA proteins. When DnaA proteins are in their ATP-bound form, they bind to the five DnaA boxes in *oriC* to initiate DNA replication. DnaA proteins also bind to each other to form a complex (**Figure 13.6**). With the aid of other DNA-binding proteins not depicted in this figure, this causes the DNA to bend around the complex of DnaA proteins and results in the separation of the AT-rich region. Because only two hydrogen bonds form between AT base pairs, whereas three hydrogen bonds occur between G and C, the DNA strands are more easily separated at an AT-rich region.

Following separation of the AT-rich region, the DnaA proteins, with the help of the DnaC protein, recruit **DNA helicase** proteins to this site. DNA helicase is also known as DnaB protein. When a DNA helicase encounters a double-stranded region, it breaks the hydrogen bonds between the two strands, thereby


5′
3′
3′
5′
AT-rich region
DnaA boxes
AT-rich region
DnaA protein
5′
3′
3′
5′
DNA helicase
5′
3′
3′
5′
Fork
Fork
5′
3′
3′
5′

DnaA proteins bind to DnaA boxes and to each other. Additional proteins that cause the DNA to bend also bind (not shown). This causes the region to wrap around the DnaA proteins and separates the AT-rich region.

DNA helicase (DnaB protein) binds to the origin. DnaC protein (not shown) assists this process.

DNA helicase separates the DNA in both directions, creating 2 replication forks.

ONLINE ANIMATION INTERACTIVE EXERCISE

FIGURE 13.6 **The events that occur at *oriC* to initiate the DNA replication process.** To initiate DNA replication, DnaA proteins bind to the five DnaA boxes, which causes the DNA strands to separate at the AT-rich region. DnaA and DnaC proteins then recruit DNA helicase (DnaB) into this region. Each DNA helicase is composed of six subunits, which form a ring around one DNA strand and travels in the 5′ to 3′ direction. As shown here, the movement of two DNA helicase proteins separates the DNA strands beyond the *oriC* region.

Concept check: How many replication forks are formed at the origin?

generating two single strands. Two DNA helicases begin strand separation within the *oriC* region and continue to separate the DNA strands beyond the origin. These proteins use the energy from ATP hydrolysis to catalyze the separation of the double-stranded parental DNA. In *E. coli,* DNA helicases bind to single-stranded DNA and travel along the DNA in a 5′ to 3′ direction to keep the replication fork moving. As shown in Figure 13.6, the action of DNA helicases promotes the movement of two replication forks outward from *oriC* in opposite directions, an event termed **bidirectional replication.**

The GATC methylation sites within oriC help to regulate the replication process. Prior to DNA replication, these sites are methylated in both strands. This full methylation of the GATC sites facilitates the initiation of DNA replication at the origin. Following DNA replication, the newly made strands are not methylated, because adenine rather than methyladenine is incorporated into the daughter strands. The initiation of DNA replication at the origin does not readily occur until after it has become fully methylated. Because it takes several minutes to methylate the GATC sequences within this region, DNA replication does not occur again too quickly.

13.2 REVIEWING THE KEY CONCEPTS

- Bacterial DNA replication begins at a single origin of replication and proceeds bidirectionally around the circular chromosome (see Figure 13.4).
- In *E. coli*, DNA replication is initiated when DnaA proteins bind to five DnaA boxes at the origin of replication and cause the AT-rich region to unwind. DNA helicases continue to separate the DNA strands, thereby promoting the movement of two replication forks (see Figures 13.5, 13.6).

13.2 COMPREHENSION QUESTIONS

1. A site in a chromosome where DNA replication begins is
   a. a promoter.
   b. an origin of replication.
   c. an operator.
   d. a replication fork.
2. The origin of replication in *E. coli* contains
   a. an AT-rich region.
   b. DnaA box sequences.
   c. GATC methylation sites.
   d. all of the above.

13.3 BACTERIAL DNA REPLICATION: SYNTHESIS OF NEW DNA STRANDS

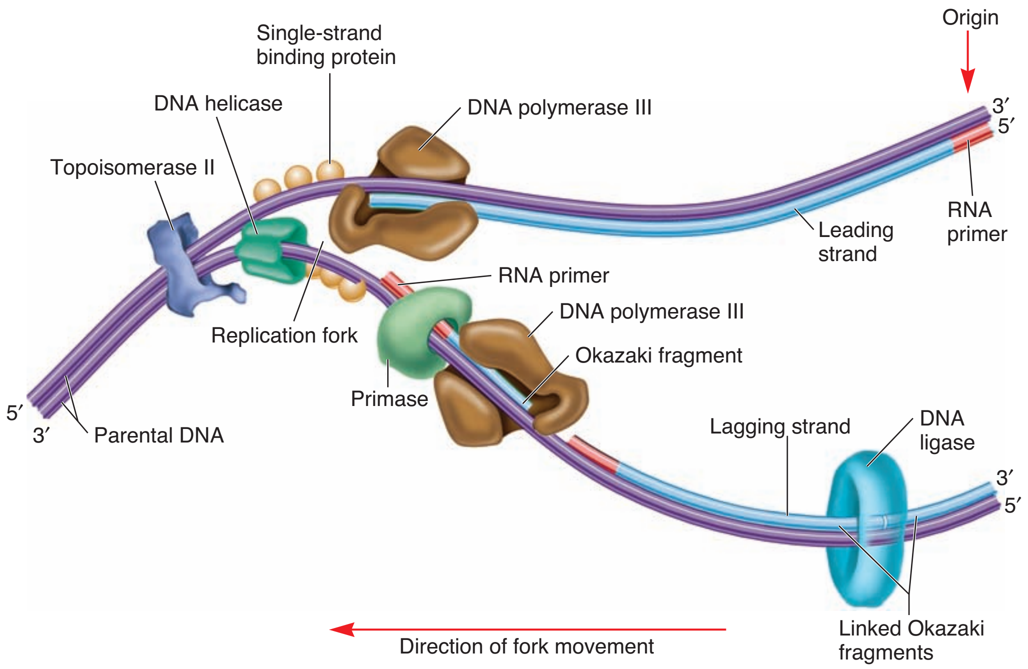
As we have seen, bacterial DNA replication begins at the origin of replication. The synthesis of new DNA strands is a stepwise process in which many cellular proteins participate. In this section, we will examine how DNA strands are made at a replication fork.

DNA Replication Occurs at the Replication Fork

Figure 13.7 provides an overview of the molecular events that occur as one of the two replication forks moves around the

Functions of key proteins involved with DNA replication

- DNA helicase breaks the hydrogen bonds between the DNA strands.
- Topoisomerase alleviates positive supercoiling.
- Single-strand binding proteins keep the parental strands apart.
- Primase synthesizes an RNA primer.
- DNA polymerase III synthesizes a daughter strand of DNA.
- DNA polymerase I excises the RNA primers and fills in with DNA (not shown).
- DNA ligase covalently links the Okazaki fragments together.



ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 13.7 **The proteins involved with DNA replication.** *Note:* The drawing of DNA polymerase III depicts the catalytic subunit that synthesizes DNA.

Concept check: *Why is primase needed for DNA replication?*

TABLE **13.1**

Proteins Involved in *E. coli* DNA Replication

Common Name	Function
DnaA protein	Binds to DnaA boxes within the origin to initiate DNA replication
DnaC protein	Aids DnaA in the recruitment of DNA helicase to the origin
DNA helicase (DnaB)	Separates double-stranded DNA
Topoisomerase	Removes positive supercoiling ahead of the replication fork
Single-strand binding protein	Binds to single-stranded DNA and prevents it from re-forming a double-stranded structure
Primase	Synthesizes short RNA primers
DNA polymerase III	Synthesizes DNA in the leading and lagging strands
DNA polymerase I	Removes RNA primers, fills in gaps with DNA
DNA ligase	Covalently attaches adjacent Okazaki fragments
Tus	Binds to ter sequences and prevents the advancement of the replication fork

bacterial chromosome. **Table 13.1** summarizes the functions of the major proteins involved in *E. coli* DNA replication.

Unwinding of the Helix Let's begin with the underwinding of the double helix. To act as a template for DNA replication, the strands of a double helix must separate. As mentioned previously, the function of DNA helicase is to break the hydrogen bonds between base pairs and thereby unwind the strands. This action generates positive supercoiling ahead of each replication fork. As shown in Figure 13.7, an enzyme known as a **topoisomerase II,** also called **DNA gyrase,** travels in front of DNA helicase and relaxes positive supercoiling.

After the two parental DNA strands have been separated and the supercoiling relaxed, they must be kept in that state until the complementary daughter strands have been made. What prevents the DNA strands from coming back together? DNA replication requires **single-strand binding proteins** that bind to the strands of parental DNA and prevent them from re-forming a double helix. In this way, the bases within the parental strands are kept in an exposed condition that enables them to hydrogen bond with individual nucleotides.

Synthesis of RNA Primers via Primase The next event in DNA replication involves the synthesis of short strands of RNA (rather than DNA) called **RNA primers.** These strands of RNA are synthesized by the linkage of ribonucleotides via an enzyme known as **primase.** This enzyme synthesizes short strands of RNA, typically 10 to 12 nucleotides in length. These short RNA strands start, or prime, the process of DNA replication. In the **leading strand,** which is made in the same direction that the replication fork is moving, a single primer is made at the origin of replication. In the **lagging strand,** which is made in short fragments in the opposite direction that the fork is moving, multiple primers are made. As discussed later, the RNA primers are eventually removed.

Synthesis of DNA via DNA Polymerase An enzyme known as **DNA polymerase** is responsible for synthesizing the DNA of the leading and lagging strands. This enzyme catalyzes the formation of covalent bonds between adjacent nucleotides, thereby making the new daughter strands. In *E. coli*, five distinct proteins function as DNA polymerases and are designated polymerase I, II, III, IV, and V. DNA polymerases I and III are involved in normal DNA replication, whereas DNA polymerases II, IV, and V play a role in DNA repair and the replication of damaged DNA.

DNA polymerase III is responsible for most of the DNA replication. It is a large enzyme consisting of 10 different subunits that play various roles in the DNA replication process (**Table 13.2**). The α subunit catalyzes the bond formation between adjacent nucleotides, and the remaining nine subunits fulfill other functions. The complex of all 10 subunits together is called DNA polymerase III holoenzyme. By comparison, DNA polymerase I is composed of a single subunit. Its role during DNA replication is to remove the RNA primers and fill in the vacant regions with DNA.

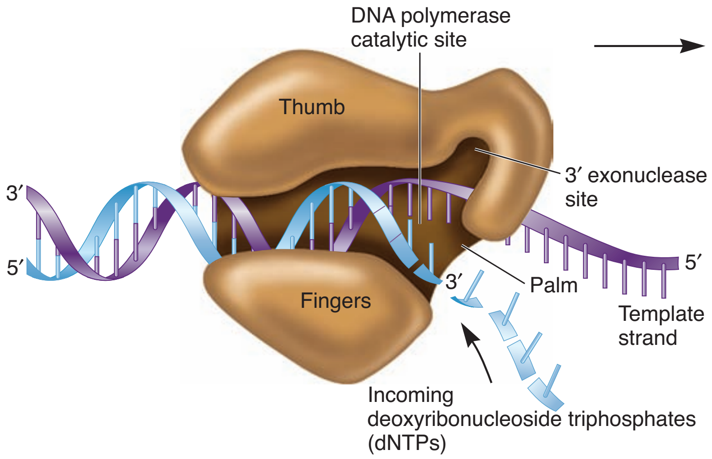
Though the various DNA polymerases in *E. coli* and other bacterial species vary in their subunit composition, several common structural features have emerged. The catalytic subunit of all DNA polymerases has a structure that resembles a human hand. As shown in **Figure 13.8**, the template DNA is threaded through the palm of the hand, with the thumb and fingers wrapped around the DNA. Incoming deoxyribonucleoside triphosphates (dNTPs) enter the catalytic site, binding to the template strand according to the AT/GC rule. As described later, deoxyribonucleoside monophosphates are covalently attached to the 3′ end of the growing strand, and pyrophosphate is released. DNA polymerase also contains a 3′ exonuclease site that removes mismatched bases, which is also described later.

As researchers began to unravel the function of DNA polymerase, two features seemed unusual (**Figure 13.9**). DNA polymerase cannot begin DNA synthesis by linking together the first two individual nucleotides. Rather, this type of enzyme can elongate only a preexisting strand starting with an RNA primer or

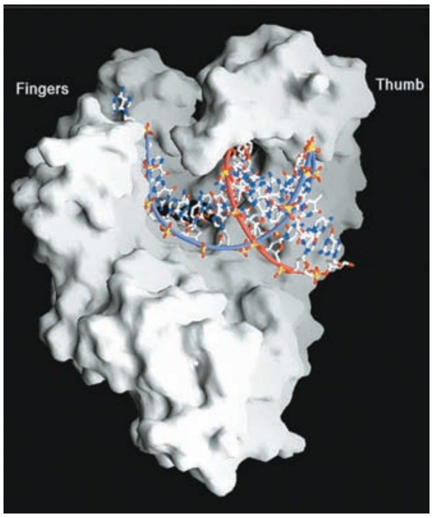
TABLE **13.2**

Subunit Composition of DNA Polymerase III Holoenzyme from *E. coli*

Subunit(s)	Function
α	Synthesizes DNA
ε	3′ to 5′ proofreading (removes mismatched nucleotides)
θ	Accessory protein that stimulates the proofreading function
β	Clamp protein, which allows DNA polymerase to slide along the DNA without falling off
τ, γ, δ, δ′, ψ, and χ	Clamp loader complex, involved with helping the clamp protein bind to the DNA


DNA polymerase catalytic site
Thumb
3′ exonuclease site
3′
5′
Fingers
3′
Palm
5′
Template strand
Incoming deoxyribonucleoside triphosphates (dNTPs)

(a) Schematic side view of DNA polymerase III


Fingers
Thumb

(b) Molecular model for DNA polymerase bound to DNA

FIGURE 13.8 **The action of DNA polymerase. (a)** DNA polymerase slides along the template strand as it synthesizes a new strand. As described later in Figure 13.15, deoxyribonucleoside triphosphates (dNTPs) initially bind to the template strand and deoxyribonucleoside monophosphates are connected in a 5′ to 3′ direction. The catalytic subunit of DNA polymerase resembles a hand that is wrapped around the template strand. In this regard, the movement of DNA polymerase along the template strand is similar to a hand that is sliding along a rope. **(b)** The molecular structure of DNA polymerase I from the bacterium *Thermus aquaticus*. This model shows a portion of DNA polymerase I that is bound to DNA. This molecular structure depicts a front view of DNA polymerase; part (a) is a schematic side view.
(Reprinted by permission from Macmillan Publishers Ltd. Ying Li, S. Korolev, G. Waksman. [1998] Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: Structural basis for nucleotide incorporation. *Embo J 17*:24, 7514–7525.)

Concept check: Is the template strand read in the 5′ to 3′ or the 3′ to 5′ direction?

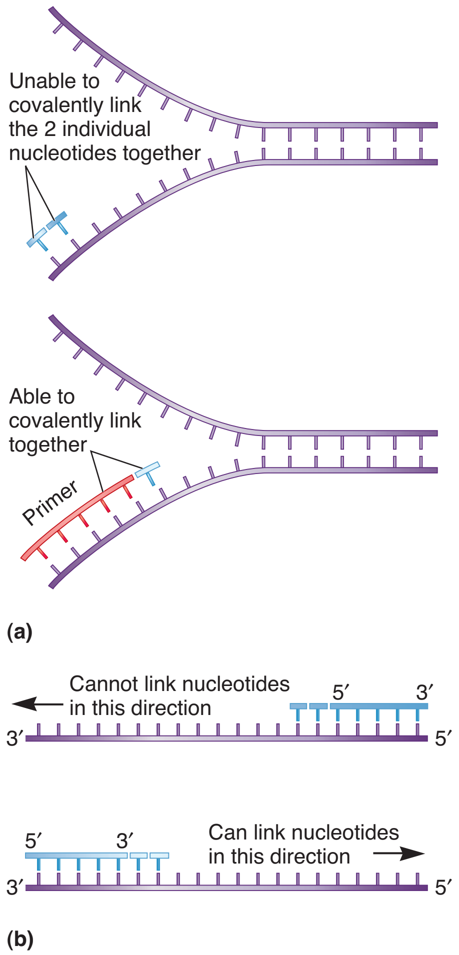

Unable to covalently link the 2 individual nucleotides together
Able to covalently link together
Primer
(a)
Cannot link nucleotides in this direction
5′
3′
3′
5′
5′
3′
Can link nucleotides in this direction
3′
5′
(b)

FIGURE 13.9 **Unusual features of DNA polymerase function. (a)** DNA polymerase can elongate a strand only from an RNA primer or an existing DNA strand. **(b)** DNA polymerase can attach nucleotides only in a 5′ to 3′ direction. *Note:* The template strand is read in the opposite, 3′ to 5′, direction.

an existing DNA strand (Figure 13.9a). A second unusual feature is the directionality of strand synthesis. DNA polymerase can attach nucleotides only in the 5′ to 3′ direction, not in the 3′ to 5′ direction (Figure 13.9b).

The Synthesis of the Leading and Lagging Strands Is Distinctly Different

Due to the unusual features of DNA replication described in Figure 13.9, the synthesis of the leading and lagging strands is distinctly different (Figure 13.10). The synthesis of RNA primers by primase allows DNA polymerase III to begin the synthesis of complementary daughter strands of DNA. DNA polymerase III catalyzes the attachment of nucleotides to the 3′ end of each primer, in a 5′ to 3′ direction. In the leading strand, one RNA primer is made at the origin, and then DNA polymerase III can attach nucleotides in a 5′ to 3′ direction as it slides toward the opening of the replication fork. The synthesis of the leading strand is therefore continuous.

In the lagging strand, the synthesis of DNA also elongates in a 5′ to 3′ manner, but it does so in the direction away from the

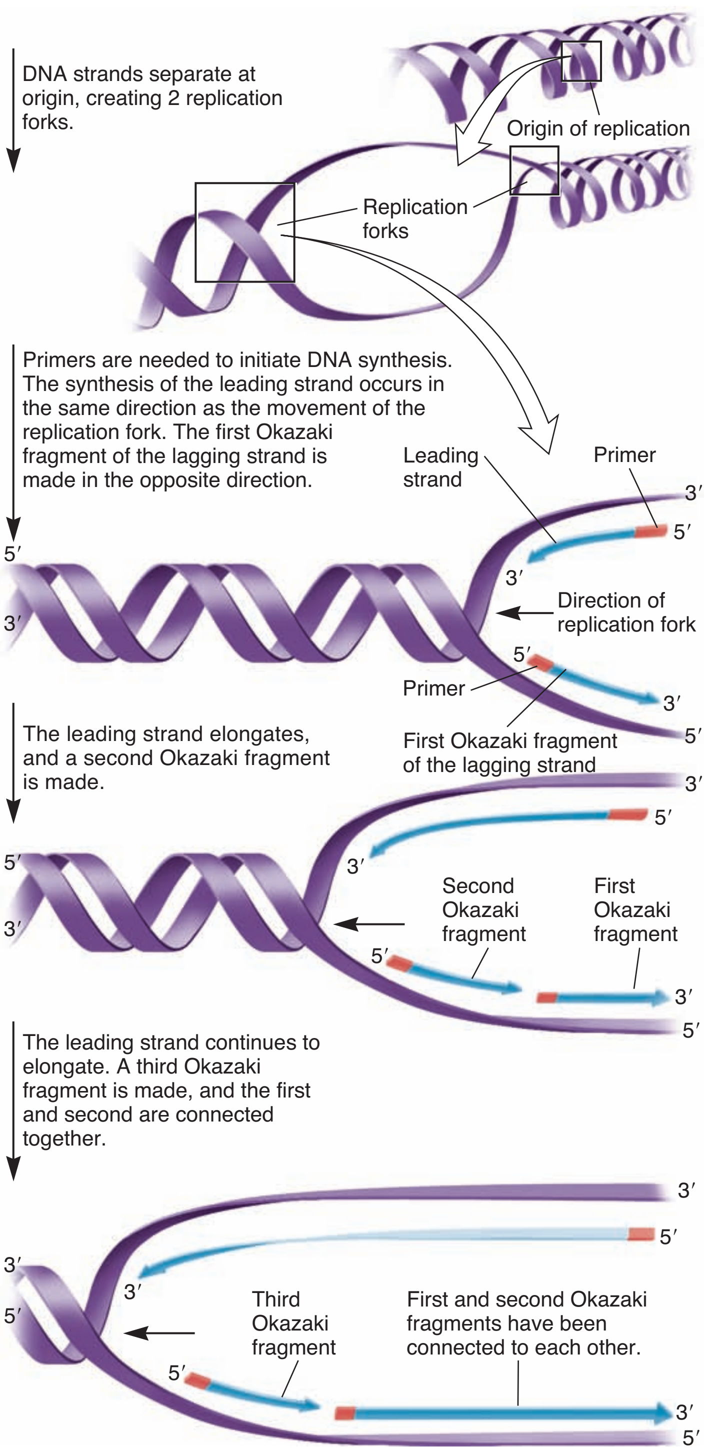

DNA strands separate at origin, creating 2 replication forks.
Origin of replication
Replication forks
Primers are needed to initiate DNA synthesis. The synthesis of the leading strand occurs in the same direction as the movement of the replication fork. The first Okazaki fragment of the lagging strand is made in the opposite direction.
Leading strand
Primer
3′
5′
5′
3′
3′
Direction of replication fork
5′
Primer
3′
5′
First Okazaki fragment of the lagging strand
The leading strand elongates, and a second Okazaki fragment is made.
3′
5′
5′
3′
3′
Second Okazaki fragment
First Okazaki fragment
5′
3′
5′
The leading strand continues to elongate. A third Okazaki fragment is made, and the first and second are connected together.
3′
5′
3′
5′
3′
Third Okazaki fragment
First and second Okazaki fragments have been connected to each other.
5′
3′
5′

FIGURE 13.10 **The synthesis of DNA at the replication fork.**

Concept check: Describe the differences between the synthesis of the leading and lagging strands.

replication fork. In the lagging strand, RNA primers must repeatedly initiate the synthesis of short segments of DNA. Therefore, the synthesis has to be discontinuous. These fragments in bacteria are typically 1000 to 2000 nucleotides long. (In eukaryotes, the fragments are shorter—100 to 200 nucleotides.) Each fragment contains a short RNA primer at the 5′ end, which is made by primase. The remainder of the fragment is a strand of DNA made by DNA polymerase III. The DNA fragments made in this manner are known as **Okazaki fragments,** after Reiji and Tuneko Okazaki, who initially discovered them in the late 1960s.

To complete the synthesis of Okazaki fragments within the lagging strand, three additional events must occur: removal of the RNA primers, synthesis of DNA in the area where the primers have been removed, and the covalent attachment of adjacent fragments of DNA (see Figure 13.10 and refer back to Figure 13.7). In *E. coli*, the RNA primers are removed by the action of DNA polymerase I. This enzyme has a 5′ to 3′ exonuclease activity, which means that DNA polymerase I digests away the RNA primers in a 5′ to 3′ direction, leaving a vacant area. DNA polymerase I then synthesizes DNA to fill in this region. It uses the 3′ end of an adjacent Okazaki fragment as a primer. For example, in Figure 13.10, DNA polymerase I would remove the RNA primer from the first Okazaki fragment and then synthesize DNA in the vacant region by attaching nucleotides to the 3′ end of the second Okazaki fragment.

After the gap has been completely filled in, a covalent bond is still missing between the last nucleotide added by DNA polymerase I and the adjacent DNA strand that had been previously made by DNA polymerase III. An enzyme known as **DNA ligase** catalyzes a covalent bond between adjacent fragments to complete the replication process in the lagging strand (refer back to Figure 13.7). In *E. coli*, DNA ligase requires NAD^+ to carry out this reaction, whereas the DNA ligases found in archaea and eukaryotes require ATP.

Now that we understand how the leading and lagging strands are made, **Figure 13.11** shows how new strands are constructed from a single origin of replication. To the left of the origin, the top strand is made continuously, whereas to the right of the origin it is made in Okazaki fragments. By comparison, the synthesis of the bottom strand is just the opposite. To the left of the origin it is made in Okazaki fragments and to the right of the origin the synthesis is continuous.

Certain Enzymes of DNA Replication Bind to Each Other to Form a Complex

Figure 13.12 provides a more three-dimensional view of the DNA replication process. DNA helicase and primase are physically bound to each other to form a complex known as a **primosome.** This complex leads the way at the replication fork. The primosome tracks along the DNA, separating the parental strands and synthesizing RNA primers at regular intervals along the lagging strand. By acting within a complex, the actions of DNA helicase and primase can be better coordinated.

The primosome is physically associated with two DNA polymerase holoenzymes to form a **replisome.** As shown in Figure 13.12, two DNA polymerase III proteins act in concert to replicate the leading and lagging strands. The term **dimeric DNA polymerase** is used to describe two DNA polymerase holoenzymes that move as a unit toward the replication fork. For this to occur, the lagging strand is looped out with respect to the DNA polymerase that synthesizes the lagging strand. This loop allows

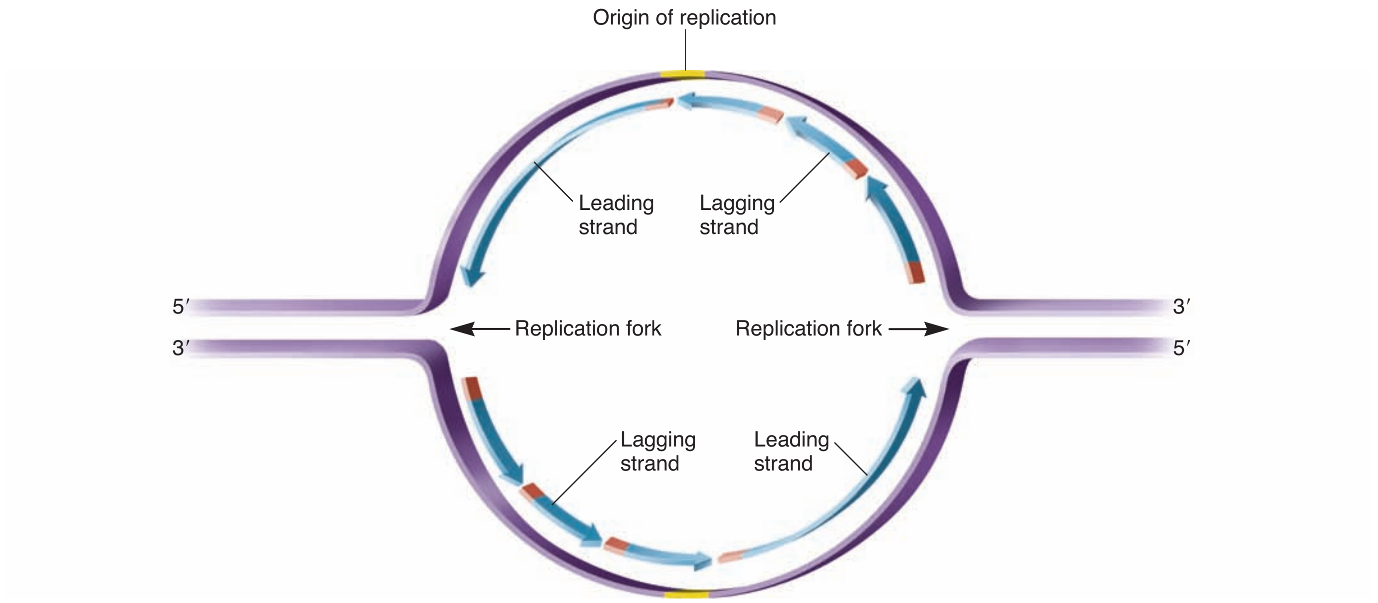

Origin of replication
Leading strand
Lagging strand
5′
3′
Replication fork
Replication fork
3′
5′
Lagging strand
Leading strand

FIGURE 13.11 **The synthesis of the leading and lagging strands outward from a single origin of replication.**

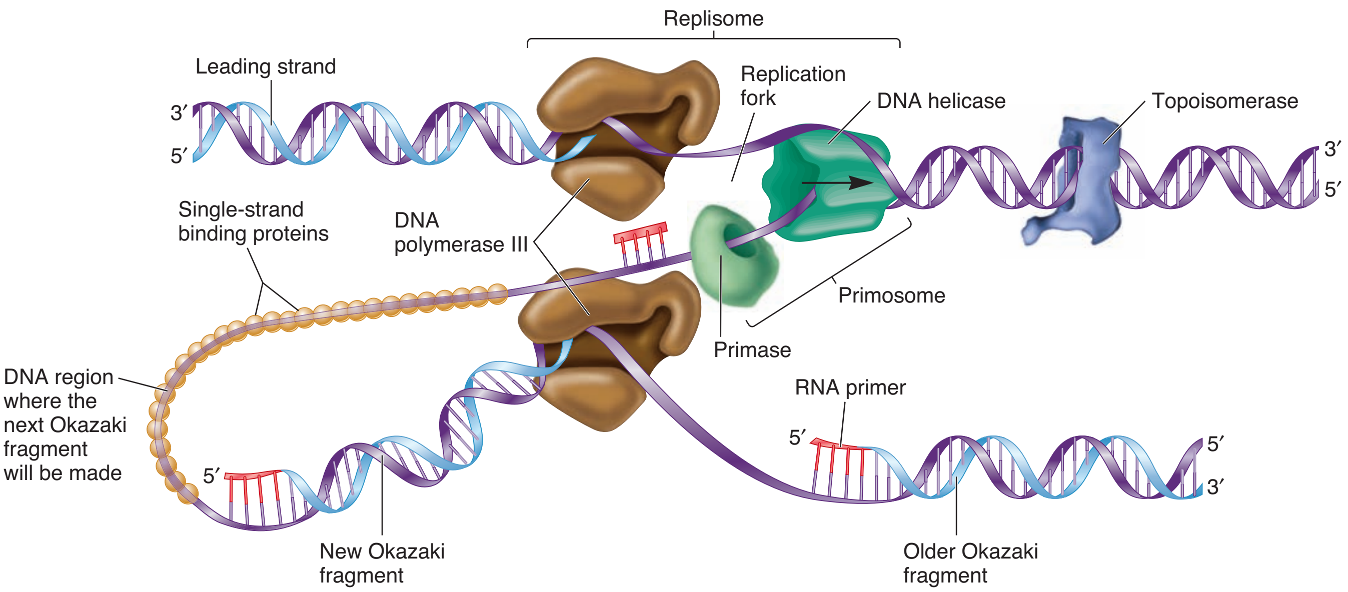

Replisome
Leading strand
3′
5′
Replication fork
DNA helicase
Topoisomerase
3′
5′
Single-strand binding proteins
DNA polymerase III
Primosome
Primase
DNA region where the next Okazaki fragment will be made
RNA primer
5′
5′
5′
3′
New Okazaki fragment
Older Okazaki fragment

FIGURE 13.12 **A three-dimensional view of DNA replication.** DNA helicase and primase associate together to form a primosome. The primosome associates with two DNA polymerase enzymes to form a replisome.

Concept check: *What is the advantage of having the replication machinery in a complex?*

the lagging-strand polymerase to make DNA in a 5′ to 3′ direction yet move toward the opening of the replication fork. Interestingly, when this DNA polymerase reaches the end of an Okazaki fragment, it must be released from the template DNA and "hop" to the RNA primer that is closest to the fork. The clamp loader complex (see Table 13.2), which is part of DNA polymerase holoenzyme, then reloads the enzyme at the site where the next RNA primer has been made. Similarly, after primase synthesizes an RNA primer in the 5′ to 3′ direction, it must hop over the primer and synthesize the next primer closer to the replication fork.

Replication Is Terminated When the Replication Forks Meet at the Termination Sequences

On the opposite side of the *E. coli* chromosome from *oriC* is a pair of **termination sequences** called ter sequences. A protein

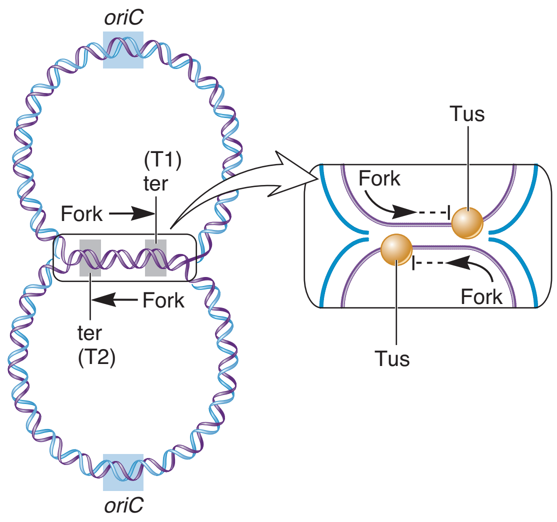


FIGURE 13.13 **The termination of DNA replication.** Two sites in the bacterial chromosome, shown as gray boxes, are ter sequences designated T1 and T2. The T1 site prevents the further advancement of the fork moving left to right, and T2 prevents the advancement of the fork moving right to left. As shown in the inset, the binding of Tus prevents the replication forks from proceeding past the ter sequences in a particular direction.

known as the termination utilization substance (Tus) binds to the ter sequences and stops the movement of the replication forks. As shown in **Figure 13.13**, one of the ter sequences designated T1 prevents the advancement of the fork moving left to right, but allows the movement of the other fork (see the inset to Figure 13.13). Alternatively, T2 prevents the advancement of the fork moving right to left, but allows the advancement of the other fork. In any given cell, only one ter sequence is required to stop the advancement of one replication fork, and then the other fork ends its synthesis of DNA when it reaches the halted replication fork. In other words, DNA replication ends when oppositely advancing forks meet, usually at T1 or T2. Finally, DNA ligase covalently links the two daughter strands, creating two circular, double-stranded molecules.

After DNA replication is completed, one last problem may exist. DNA replication often results in two intertwined DNA molecules known as **catenanes** (**Figure 13.14**). Fortunately, catenanes are only transient structures in DNA replication. In *E. coli*, topoisomerase II introduces a temporary break into the DNA strands and then rejoins them after the strands have become unlocked. This allows the catenanes to be separated into individual circular molecules.

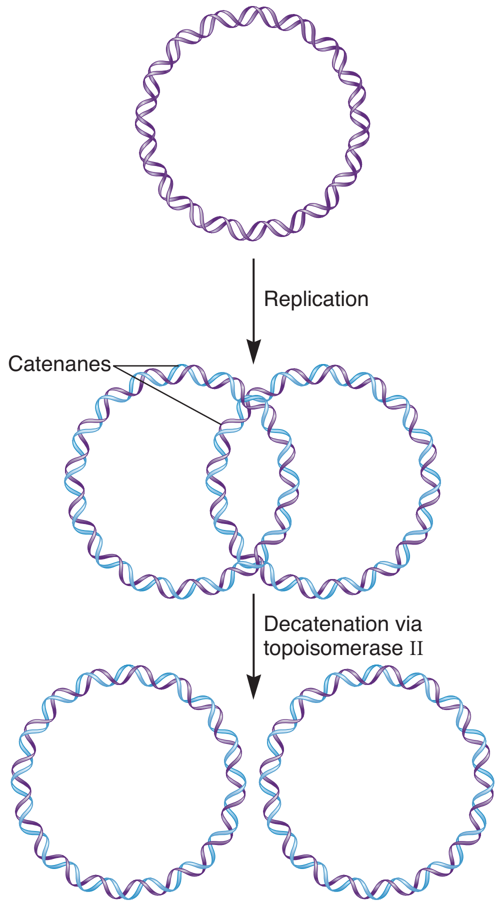


FIGURE 13.14 **Separation of catenanes.** DNA replication can result in two intertwined chromosomes called catenanes. These catenanes can be separated by the action of topoisomerase II.

Concept check: Is DNA strand breakage needed for catenane separation?

13.3 REVIEWING THE KEY CONCEPTS

- At each replication fork, DNA helicase unwinds the DNA and topisomerase alleviates positive supercoiling. Single-strand binding proteins coat the DNA to prevent the strands from coming back together. Primase synthesizes RNA primers and DNA polymerase synthesizes complementary strands of DNA. DNA ligase seals the gaps between Okazaki fragments (see Figure 13.7, Table 13.1).
- DNA polymerase III in *E. coli* is an enzyme with several subunits. The catalytic subunit wraps around the DNA like a hand (see Figure 13.8, Table 13.2).
- DNA polymerase needs a primer to synthesize DNA and makes new DNA strands in a 5′ to 3′ direction (see Figure 13.9).
- During DNA synthesis, the leading strand is made continuously in the direction of the replication fork, whereas the lagging strand is made as Okazaki fragments in the direction away from the fork (see Figures 13.10, 13.11).
- The primosome is a complex between helicase and primase. The replisome is a complex between the primosome and dimeric DNA polymerase (see Figure 13.12).
- In *E. coli*, DNA replication is terminated at ter sequences (see Figure 13.13).
- Following DNA replication, interlocked catenanes sometimes need to be unlocked via topoisomerase II (see Figure 13.14).

13.3 COMPREHENSION QUESTIONS

1. The enzyme known as _________ uses _________ and separates the DNA strands at the replication fork.
   a. helicase, ATP
   b. helicase, GTP
   c. gyrase, ATP
   d. gyrase, GTP
2. In the lagging strand, DNA is made in the direction __________ the replication fork and is made as ___________.
   a. toward, one continuous strand
   b. away from, one continuous strand
   c. toward, Okazaki fragments
   d. away from, Okazaki fragments

13.4 BACTERIAL DNA REPLICATION: CHEMISTRY AND ACCURACY

In Sections 13.2 and 13.3, we examined the origin of replication and considered how DNA strands are synthesized. In this section, we will take a closer look at the process in which nucleotides are attached to a growing DNA strand and discuss how DNA replication is amazingly accurate.

DNA Polymerase III Is a Processive Enzyme That Uses Deoxyribonucleoside Triphosphates

As shown in **Figure 13.15**, the nucleotide about to be attached to the growing strand is a dNTP containing three phosphate groups attached at the 5′–carbon atom of deoxyribose. The dNTP binds to the template strand according to the AT/GC rule. Next, the 3′–OH group on the previous nucleotide reacts with the phosphate group adjacent to the sugar on the incoming nucleotide. The breakage of a covalent bond between two phosphates in a dNTP is a highly exergonic reaction that provides the energy to form a covalent (ester) bond between the sugar at the 3′ end of the DNA strand and the phosphate of the incoming nucleotide. The formation of this covalent bond causes the newly made strand to grow in the 5′ to 3′ direction. Pyrophosphate (PP_i) is released.

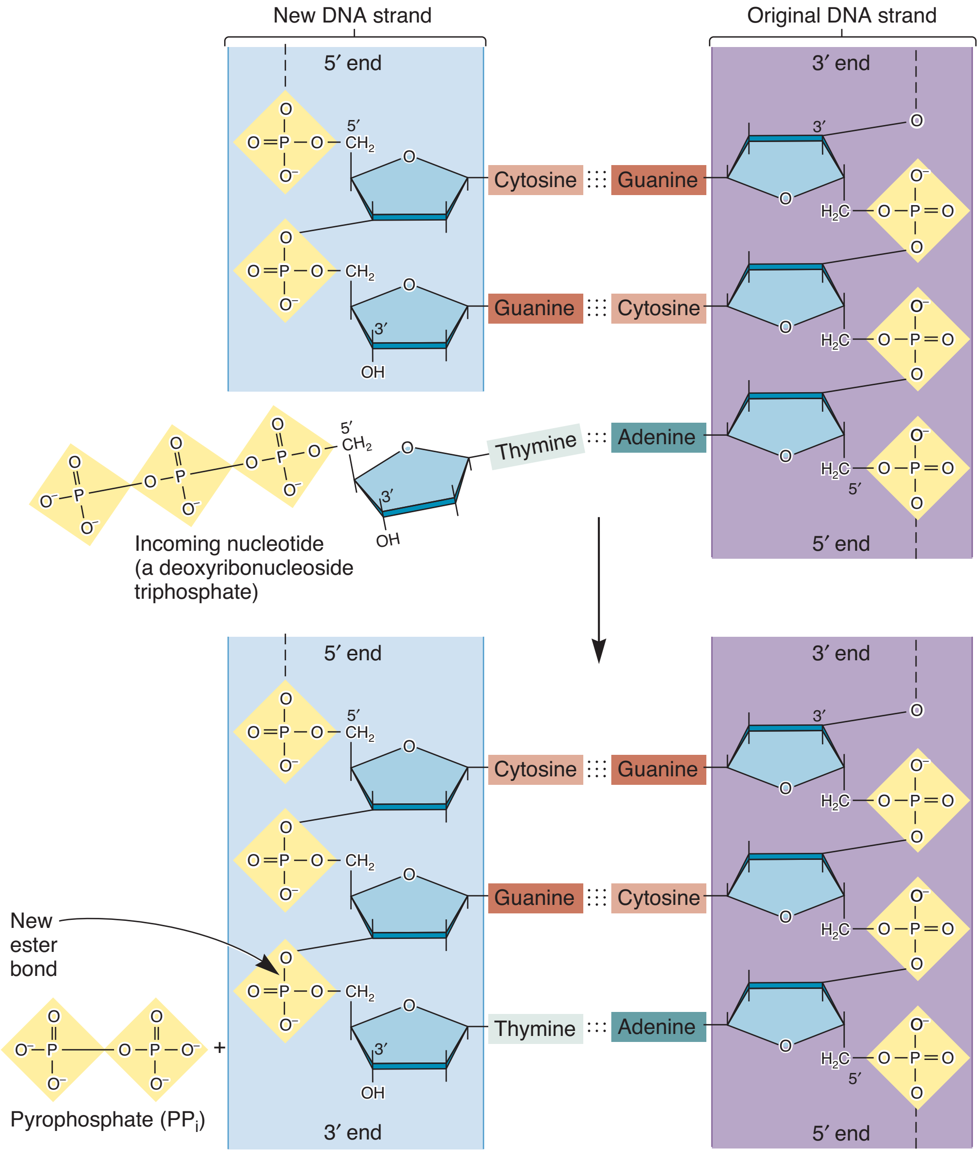


FIGURE 13.15 The attachment of a nucleotide to a DNA strand. An incoming deoxyribonucleoside triphosphate (dNTP) is cleaved to form a deoxyribonucleoside monophosphate and pyrophosphate (PP_i). The energy released from this exergonic reaction allows the deoxyribonucleoside monophosphate to form a covalent (ester) bond at the 3′ end of the growing strand. This reaction is catalyzed by DNA polymerase. PP_i is released.

Concept check: Does the oxygen in the newly made ester bond come from the phosphate or from the sugar?

As noted in Chapter 11 (Figure 11.10), the term *phosphodiester linkage* is used to describe the linkage between a phosphate and two sugar molecules. As its name implies, a phosphodiester linkage involves two ester bonds. In comparison, as a DNA strand grows, a single covalent (ester) bond is formed between adjacent nucleotides (see Figure 13.15). The other ester bond in the phosphodiester linkage—the bond between the 5′-oxygen and phosphorus—is already present in the incoming nucleotide.

DNA polymerase catalyzes the covalent attachment of nucleotides with great speed. In *E. coli*, DNA polymerase III attaches approximately 750 nucleotides per second! DNA polymerase III can catalyze the synthesis of the daughter strands so quickly because it is a **processive enzyme.** This means it does not dissociate from the growing strand after it has catalyzed the covalent joining of two nucleotides. Rather, as depicted in Figure 13.8a, it remains clamped to the DNA template strand and slides along the template as it catalyzes the synthesis of the daughter strand. The β subunit of the holoenzyme, also known as the clamp protein, promotes the association of the holoenzyme with the DNA as it glides along the template strand (refer back to Table 13.2). The β subunit forms a dimer in the shape of a ring. The hole of the ring is large enough to accommodate a double-stranded DNA molecule, and its width is about one turn of DNA. A complex of several subunits functions as a clamp loader that allows the DNA polymerase holoenzyme to initially clamp onto the DNA.

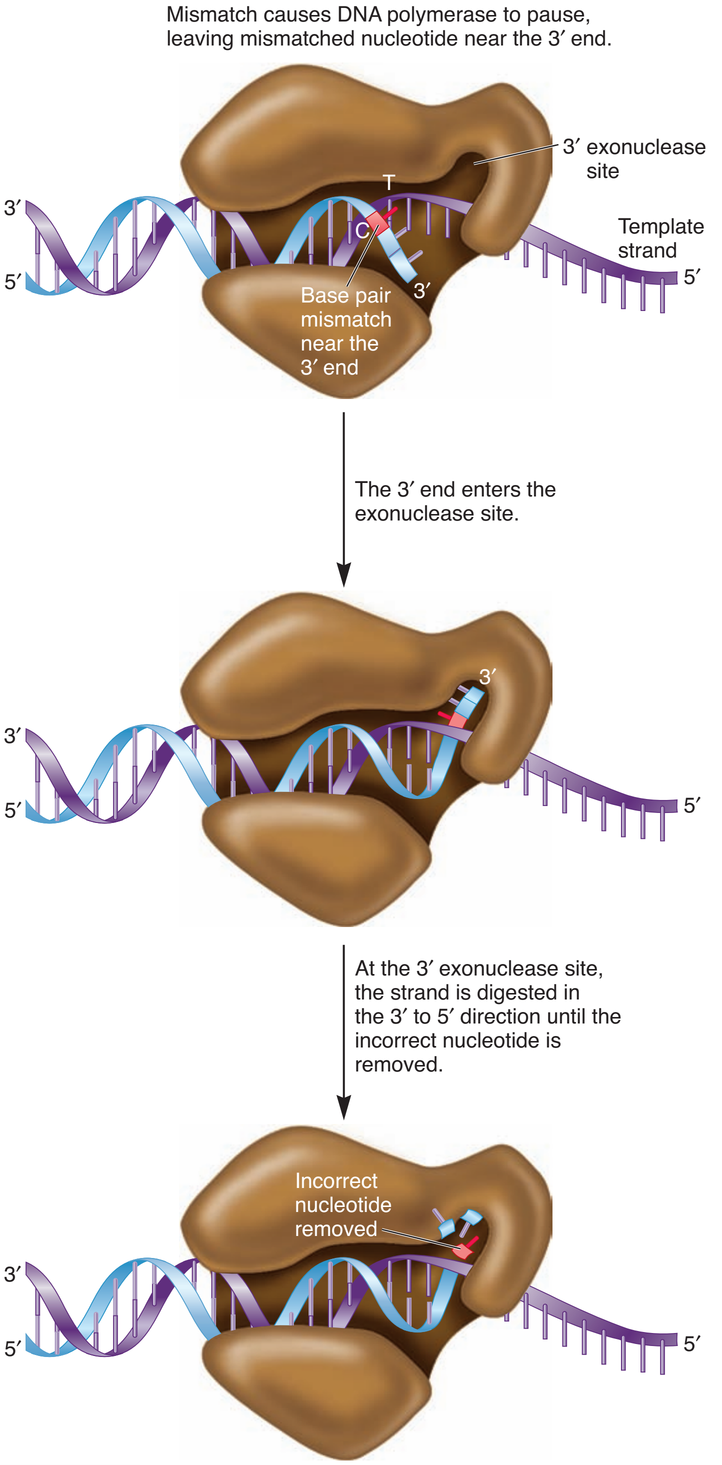
The effects of processivity are really quite remarkable. In the absence of the β subunit, DNA polymerase can synthesize DNA at a rate of approximately 20 nucleotides per second. On average, it falls off the DNA template after about 10 nucleotides have been linked together. By comparison, when the β subunit is present, as in the holoenzyme, the synthesis rate is approximately 750 nucleotides per second. In the leading strand, DNA polymerase III has been estimated to synthesize a segment of DNA that is over 500,000 nucleotides in length before it inadvertently falls off.

DNA Replication Is Very Accurate

With replication occurring so rapidly, one might imagine that mistakes could happen in which the wrong nucleotide is incorporated into the growing daughter strand. Although mistakes do happen during DNA replication, they are extraordinarily rare. In the case of DNA synthesis via DNA polymerase III, only one mistake per 100 million nucleotides is made. Therefore, DNA synthesis occurs with a high degree of accuracy, or **fidelity.**

Why is the fidelity so high? First, the hydrogen bonding between G and C or A and T is much more stable than between mismatched pairs. However, this stability accounts for only part of the fidelity, because mismatching due to stability considerations would account for 1 mistake per 1000 nucleotides.

Two characteristics of DNA polymerase also contribute to the fidelity of DNA replication. First, the active site of DNA polymerase preferentially catalyzes the attachment of nucleotides when the correct bases are located in opposite strands. Helix distortions caused by mispairing usually prevent an incorrect nucleotide from properly occupying the active site of DNA polymerase. By comparison, the active site undergoes a change in shape when the correct nucleotide binds, a process called induced fit, which is necessary for catalysis. The ability of correct nucleotides to cause induced fit decreases the error rate to a range of 1 in 100,000 to 1 million.


Mismatch causes DNA polymerase to pause, leaving mismatched nucleotide near the 3′ end.
3′
5′
T
C
3′ exonuclease site
Template strand
5′
3′
Base pair mismatch near the 3′ end
The 3′ end enters the exonuclease site.
3′
5′
3′
5′
At the 3′ exonuclease site, the strand is digested in the 3′ to 5′ direction until the incorrect nucleotide is removed.
3′
5′
Incorrect nucleotide removed
5′

ONLINE ANIMATION

FIGURE 13.16 **The proofreading function of DNA polymerase.** When a base pair mismatch is found, the end of the newly made strand is shifted into the 3′ exonuclease site. The DNA is digested in the 3′ to 5′ direction to release the incorrect nucleotide.

A second way that DNA polymerase decreases the error rate is by the enzymatic removal of mismatched nucleotides. As shown in **Figure 13.16**, DNA polymerase can identify a mismatched nucleotide and remove it from the daughter strand. This occurs by exonuclease cleavage of the bonds between adjacent nucleotides at the 3′ end of the newly made strand. The ability of DNA polymerase to remove mismatched bases by this mechanism is called **proofreading.** Proofreading occurs by the removal of nucleotides in the 3′ to 5′ direction at the 3′ exonuclease site. After the mismatched nucleotide is removed, DNA polymerase resumes DNA synthesis in the 5′ to 3′ direction.

13.4 REVIEWING THE KEY CONCEPTS

- DNA polymerase III is a processive enzyme that uses deoxyribonucleoside triphosphates (dNTPs) to make new DNA strands (see Figure 13.15).
- The high fidelity of DNA replication is a result of (1) the stability of hydrogen bonding between the correct bases, (2) the phenomenon of induced fit, and (3) the proofreading ability of DNA polymerase (see Figure 13.16).

13.4 COMPREHENSION QUESTIONS

1. DNA polymerase III is a processive enzyme, which means that
   a. it does not dissociate from the growing strand after it has attached a nucleotide to the 3′ end.
   b. it makes a new strand very quickly.
   c. it proceeds toward the opening of the replication fork.
   d. it copies DNA with relatively few errors.
2. The proofreading function of DNA polymerase involves the recognition of a ________ and the removal of a short segment of DNA in the __________ direction.
   a. missing base, 5′ to 3′
   b. base mismatch, 5′ to 3′
   c. missing base, 3′ to 5′
   d. base mismatch, 3′ to 5′

13.5 EUKARYOTIC DNA REPLICATION

Eukaryotic DNA replication is not as well understood as bacterial replication. Much research has been carried out on a variety of experimental organisms, particularly yeast and mammalian cells. Many of these studies have found extensive similarities between the general features of DNA replication in prokaryotes and eukaryotes. For example, DNA helicases, topoisomerases, single-strand binding proteins, primases, DNA polymerases, and DNA ligases—the types of bacterial proteins described in Table 13.1—have also been identified in eukaryotes. Nevertheless, at the molecular level, eukaryotic DNA replication appears to be substantially more complex. These additional intricacies of eukaryotic DNA replication are related to several features of eukaryotic cells. In particular, eukaryotic cells have larger, linear chromosomes with multiple origins of replication; the chromatin is tightly packed within nucleosomes; and cell-cycle regulation is much more complicated. This section emphasizes some of the unique features of eukaryotic DNA replication.

Initiation Occurs at Multiple Origins of Replication on Linear Eukaryotic Chromosomes

Eukaryotes have long, linear chromosomes with multiple origins of replication, which enable the DNA to be replicated in a reasonable length of time. In 1968, Joel Huberman and Arthur Riggs provided evidence for multiple origins of replication by adding a pulse of a radiolabeled deoxyribonucleoside (^{3}H-thymidine) to a culture of actively dividing cells, followed by a chase with nonlabeled thymidine. In this pulse/chase experiment, the radiolabeled thymidine was taken up by the cells and incorporated into their newly made DNA strands for a brief period. The chromosomes were then isolated from the cells and subjected to autoradiography. As seen in **Figure 13.17**, radiolabeled segments were interspersed among nonlabeled segments. This result is consistent with the hypothesis that eukaryotic chromosomes contain multiple origins of replication.

As shown schematically in **Figure 13.18**, DNA replication proceeds bidirectionally from many origins of replication during S phase of the cell cycle. The multiple replication forks eventually make contact with each other to complete the replication process.

The molecular features of some eukaryotic origins of replication may have some similarities to the origins found in bacteria. At the molecular level, eukaryotic origins of replication have been extensively studied in the yeast *Saccharomyces cerevisiae*. They have been named **ARS elements** (for autonomously replicating sequence). ARS elements, which are about 50 bp in length, are necessary to initiate chromosome replication. ARS elements have unique features of their DNA sequences. First, they contain a higher percentage of A and T bases than the rest of the chromosomal DNA. In addition, they contain a copy of the ARS consensus sequence, ATTTAT(A or G)TTTA, along with additional elements that enhance origin function. In animals, the critical features that define origins of replication are not completely understood. In many species, origins are not determined by particular DNA sequences. Rather, origins of replication in animals may be due to specific features of chromatin structure, including protein modifications.

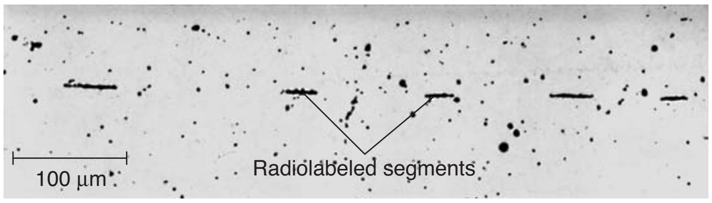


FIGURE 13.17 **Evidence for multiple origins of replication in eukaryotic chromosomes.** In this experiment, cells were given a pulse/chase of ^{3}H-thymidine and unlabeled thymidine. The chromosomes were isolated and subjected to autoradiography. In this micrograph, radiolabeled segments were interspersed among nonlabeled segments, indicating that eukaryotic chromosomes contain multiple origins of replication.

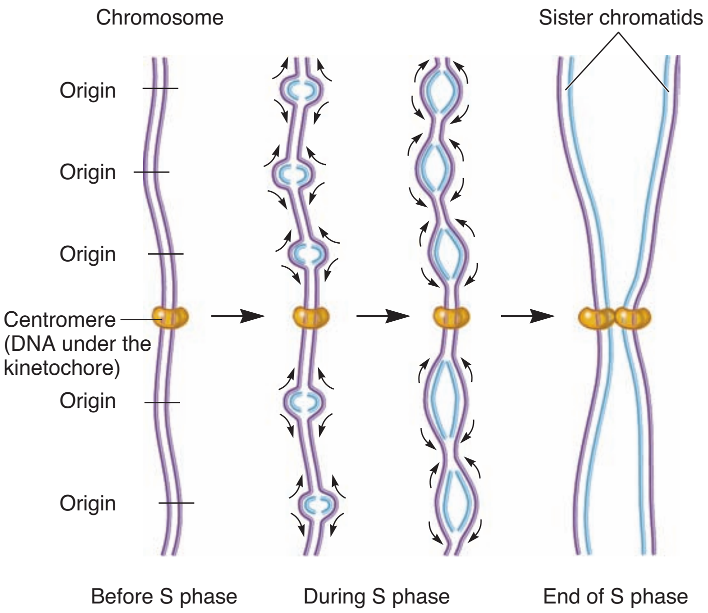


FIGURE 13.18 **The replication of eukaryotic chromosomes.** At the beginning of the S phase of the cell cycle, eukaryotic chromosome replication begins from multiple origins of replication. As S phase continues, the replication forks move bidirectionally to replicate the DNA. By the end of S phase, all of the replication forks have merged. The net result is two sister chromatids attached to each other at the centromere.

Concept check: Why do eukaryotes need multiple origins of replication?

DNA replication in eukaryotes begins with the assembly of a **prereplication complex (preRC)** consisting of at least 14 different proteins. Part of the preRC is a group of six proteins called the **origin recognition complex (ORC)** that acts as the initiator of eukaryotic DNA replication. The ORC was originally identified in yeast as a protein complex that binds directly to ARS elements. DNA replication at the origin begins with the binding of ORC, which usually occurs during G_1 phase. Other proteins of the preRC then bind, including a group of proteins called **MCM helicase.**[1] The binding of MCM helicase at the origin completes a process called **DNA replication licensing;** only those origins with MCM helicase can initiate DNA synthesis. During S phase, DNA synthesis begins when preRCs are acted on by at least 22 additional proteins that activate MCM helicase and assemble two divergent replication forks at each replication origin. An important role of these additional proteins is to carefully regulate the initiation of DNA replication so it happens at the correct time and only once at each origin of replication during the cell cycle. The precise roles of these proteins are under active research investigation.

Eukaryotes Contain Several Different DNA Polymerases

Eukaryotes have many types of DNA polymerases. For example, mammalian cells have well over a dozen different DNA polymerases (Table 13.3). Four of these, designated α (alpha), ε (epsilon), δ (delta), and γ (gamma), have the primary function of replicating DNA. DNA polymerase γ functions in the mitochondria to replicate mitochondrial DNA, whereas α, ε, and δ are involved with DNA replication in the cell nucleus during S phase.

DNA polymerase α is the only eukaryotic polymerase that associates with primase. The functional role of the DNA polymerase α/primase complex is to synthesize a short RNA–DNA primer of approximately 10 RNA nucleotides followed by 20 to 30 DNA nucleotides. This short RNA–DNA strand is then used by DNA polymerase ε or δ for the processive elongation of the leading and lagging strands, respectively. For this to happen, the DNA polymerase α/primase complex dissociates from the replication fork and is exchanged for DNA polymerase ε or δ. This exchange is called a polymerase switch**.** Accumulating evidence suggests that DNA polymerase ε is primarily involved with leading-strand synthesis, whereas DNA polymerase δ is responsible for lagging-strand synthesis.

What are the functions of the other DNA polymerases? Several of them play an important role in DNA repair, a topic that will be examined in Chapter 18. DNA polymerase β, which has been studied for several decades, is not involved in the replication of normal DNA, but plays an important role in removing incorrect bases from damaged DNA. More recently, several additional DNA polymerases have been identified. Although their precise roles have not been elucidated, many of these are in a category called **lesion-replicating polymerases.** When DNA polymerase α, δ, and ε encounter abnormalities in DNA structure, such as

TABLE **13.3**

Eukaryotic DNA Polymerases

Polymerase Types*	Function
α	Initiates DNA replication in conjunction with primase
ε	Replication of the leading strand during S phase
δ	Replication of the lagging strand during S phase
γ	Replication of mitochondrial DNA
η, κ, ι, ξ (lesion-replicating polymerases)	Replication of damaged DNA
α, β, δ, ε, σ, λ, μ, φ, θ, η	DNA repair or other functions†

*The designations are those of mammalian enzymes.

†Many DNA polymerases have dual functions. For example, DNA polymerases α, δ, and ε are involved in the replication of normal DNA and also play a role in DNA repair. In cells of the immune system, certain genes that encode antibodies (i.e., immunoglobulin genes) undergo a phenomenon known as hypermutation. This increases the variation in the kinds of antibodies the cells can make. Certain polymerases in this list, such as η, may play a role in hypermutation of immunoglobulin genes. DNA polymerase σ may play a role in sister chromatid cohesion, a topic discussed in Chapter 12.

---

[1] MCM is an abbreviation for minichromosome maintenance. The genes encoding MCM proteins were originally identified in mutant yeast strains that are defective in the maintenance of minichromosomes in the cell. MCM proteins have since been shown to play a role in DNA replication.

abnormal bases or covalent crosslinks, they may be unable to replicate over the aberration. When this occurs, lesion-replicating polymerases are attracted to the damaged DNA and have special properties that enable them to synthesize a complementary strand over the abnormal region. Each type of lesion-replicating polymerase may be able to replicate over a different kind of DNA damage.

Flap Endonuclease Removes RNA Primers During Eukaryotic DNA Replication

Another key difference between bacterial and eukaryotic DNA replication is the way that RNA primers are removed. As discussed earlier in this chapter, bacterial RNA primers are removed by DNA polymerase I. By comparison, a DNA polymerase enzyme does not play this role in eukaryotes. Instead, an enzyme called **flap endonuclease** is primarily responsible for RNA primer removal.

Flap endonuclease gets its name because it removes small pieces of RNA flaps that are generated by the action of DNA polymerase δ. As shown in **Figure 13.19**, DNA polymerase δ elongates the left Okazaki fragment until it runs into the RNA primer of the adjacent Okazaki fragment on the right. This causes a portion of the RNA primer to form a short flap, which is removed by flap endonuclease. As DNA polymerase δ continues to elongate the DNA, short flaps continue to be generated, which are sequentially removed by flap endonuclease. Eventually, all of the RNA primer is removed, and DNA ligase can seal the DNA fragments together.

Though flap endonuclease is thought to be the primary pathway for RNA primer removal in eukaryotes, it is unable to remove a flap that is too long. In such cases, a long flap is cleaved by the enzyme called Dna2 nuclease/helicase. This enzyme can cut a long flap, thereby generating a short flap that is then removed via flap endonuclease.

The Ends of Eukaryotic Chromosomes Are Replicated by Telomerase

Linear eukaryotic chromosomes contain **telomeres** at both ends. The term *telomere* refers to the telomeric sequences within the DNA and the special proteins that are bound to these sequences. Telomeric sequences consist of a moderately repetitive tandem array and a 3′ overhang region that is 12 to 16 nucleotides in length (**Figure 13.20**).

The tandem array that occurs within the telomere has been studied in a wide variety of eukaryotic organisms. A common feature is that the telomeric sequence contains several guanine nucleotides and often many thymine nucleotides (**Table 13.4**). Depending on the species and the cell type, this sequence can be tandemly repeated up to several hundred times in the telomere region.

One reason why telomeric repeat sequences are needed is because DNA polymerase is unable to replicate the 3′ ends of DNA strands. Why is DNA polymerase unable to replicate this region? The answer lies in the two unusual enzymatic features of

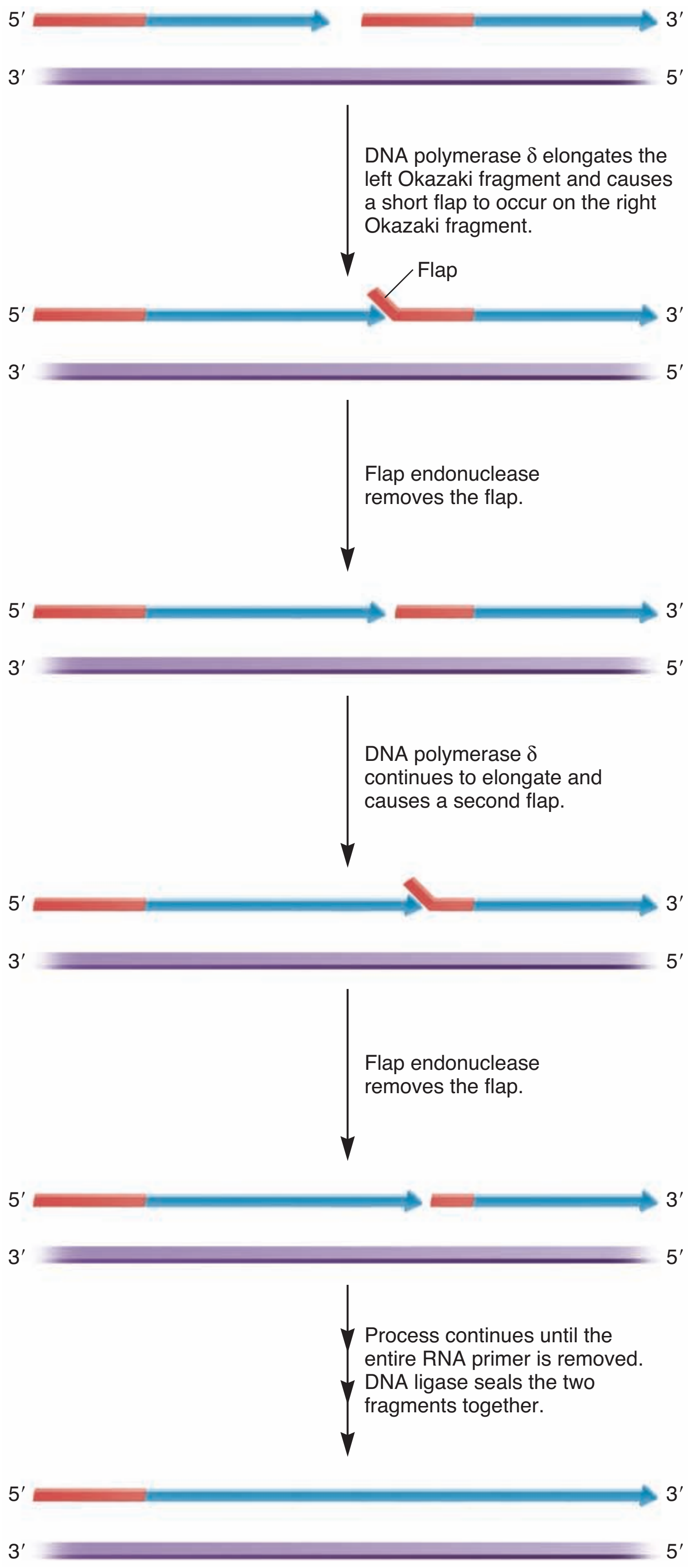

5′
3′
3′
5′
DNA polymerase δ elongates the left Okazaki fragment and causes a short flap to occur on the right Okazaki fragment.
Flap
5′
3′
3′
5′
Flap endonuclease removes the flap.
5′
3′
3′
5′
DNA polymerase δ continues to elongate and causes a second flap.
5′
3′
3′
5′
Flap endonuclease removes the flap.
5′
3′
3′
5′
Process continues until the entire RNA primer is removed. DNA ligase seals the two fragments together.
5′
3′
3′
5′

FIGURE 13.19 **Removal of an RNA primer by flap endonuclease.**

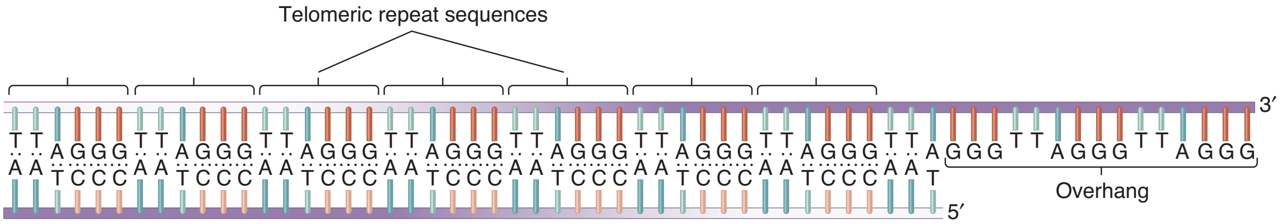

Telomeric repeat sequences
3′
TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
AATCCCAATCCCAATCCCAATCCCAATCCCAATCCCAATCCCAAT
5′
Overhang

FIGURE 13.20 **General structure of telomeric sequences.** The telomere DNA consists of a tandemly repeated sequence with a 12- to 16-nucleotide overhang.

ONLINE ANIMATION

TABLE 13.4
Telomeric Repeat Sequences Within Selected Organisms

Group	Example	Telomeric Repeat Sequence
Mammals	Humans	TTAGGG
Slime molds	*Physarum, Didymium*	TTAGGG
	Dictyostelium	$AG_{(1-8)}$
Filamentous fungi	*Neurospora*	TTAGGG
Budding yeast	*Saccharomyces cerevisiae*	$TG_{(1-3)}$
Ciliates	*Tetrahymena*	TTGGGG
	Paramecium	TTGGG(T/G)
	Euplotes	TTTTGGGG
Flowering plants	*Arabidopsis*	TTTAGGG

this enzyme. As discussed previously, DNA polymerase synthesizes DNA only in a 5′ to 3′ direction, and it cannot link together the first two individual nucleotides; it can elongate only preexisting strands. These two features of DNA polymerase function pose a problem at the 3′ ends of linear chromosomes. As shown in **Figure 13.21**, the 3′ end of a DNA strand cannot be replicated by DNA polymerase because a primer cannot be made upstream from this point. Therefore, if this problem were not solved, the chromosome would become progressively shorter with each round of DNA replication.

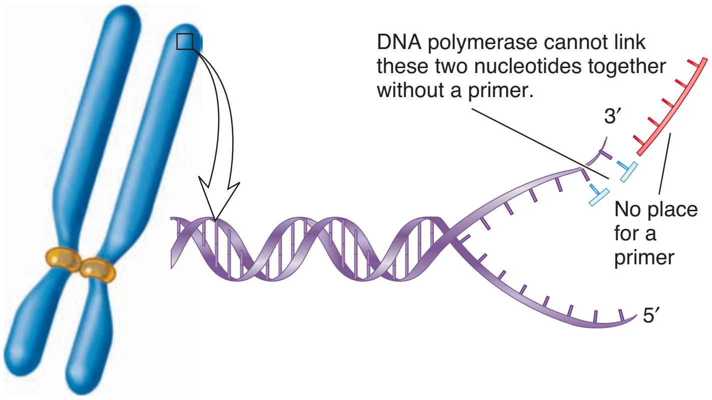
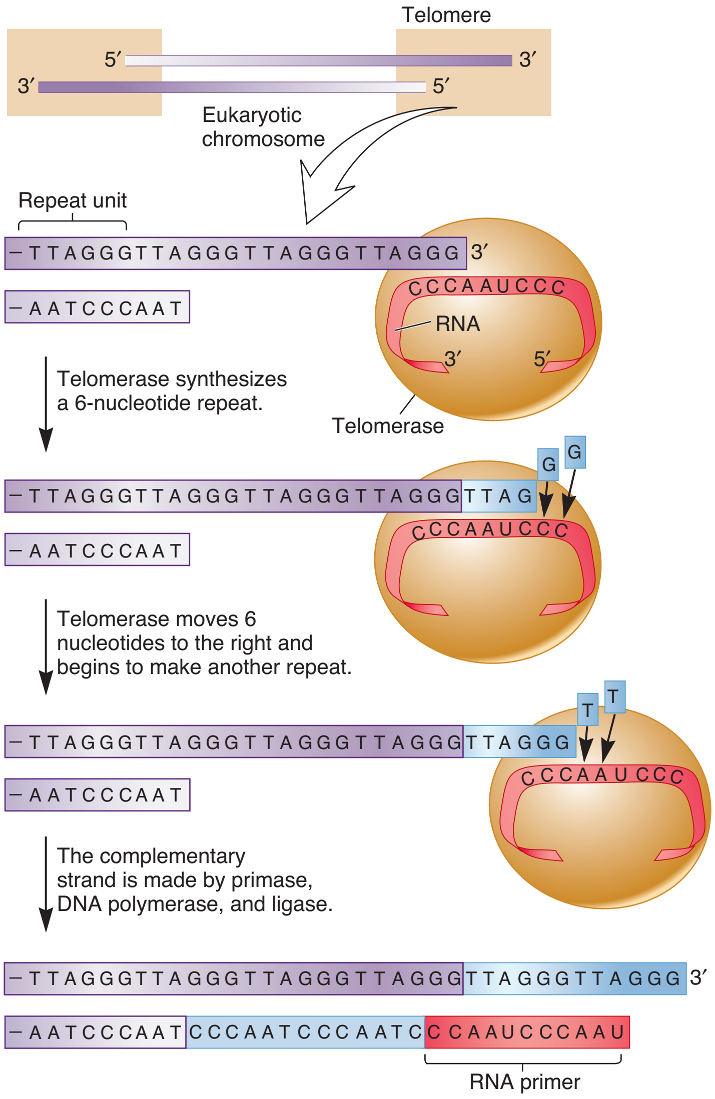

DNA polymerase cannot link these two nucleotides together without a primer.
3′
No place for a primer
5′

FIGURE 13.21 **The replication problem at the ends of linear chromosomes.** DNA polymerase cannot synthesize a DNA strand that is complementary to the 3′ end because a primer cannot be made upstream from this site.

To prevent the loss of genetic information due to chromosome shortening, additional DNA sequences are attached to the ends of telomeres. In 1984, Carol Greider and Elizabeth Blackburn discovered an enzyme called **telomerase** that prevents chromosome shortening. They received the 2009 Nobel Prize in physiology or medicine for their discovery. Telomerase recognizes the sequences at the ends of eukaryotic chromosomes and synthesizes additional repeats of telomeric sequences. **Figure 13.22** shows the interesting mechanism by which telomerase works. The telomerase enzyme contains both protein subunits and RNA. The RNA part of telomerase contains a sequence complementary to the DNA sequence found in the telomeric repeat. This allows telomerase to bind to the 3′ overhang region of the telomere. Following binding, the RNA sequence beyond the binding site functions as a template allowing the synthesis of a six-nucleotide sequence at the end of the DNA strand. This is called polymerization, because it is analogous to the function of DNA polymerase. It is catalyzed by two identical protein subunits called **telomerase reverse transcriptase (TERT).** TERT's name indicates that it uses an RNA template to synthesize DNA. Following polymerization, the telomerase can then move to the new end of this DNA strand, a process called translocation. This binding–polymerization–translocation cycle occurs many times in a row, thereby greatly lengthening the 3′ end of the DNA strand in the telomeric region. The complementary strand is then synthesized by primase, DNA polymerase, and DNA ligase, as described earlier in this chapter.

13.5 REVIEWING THE KEY CONCEPTS

- Eukaryotic chromosomes contain multiple origins of replication. Part of the prereplication complex is formed from a group of six proteins called the origin recognition complex (ORC). The binding of MCM helicase completes a process called DNA replication licensing (see Figures 13.17, 13.18).
- Eukaryotes have several different DNA polymerases with specialized roles. Different types of DNA polymerases switch with each other during the process of DNA replication (see Table 13.3).
- Flap endonuclease is an enzyme that removes RNA primers from Okazaki fragments (see Figure 13.19).
- The ends of eukaryotic chromosomes contain telomeres, which are composed of short repeat sequences and proteins (see Figure 13.20, Table 13.4).


Telomere
5′
3′
3′
5′
Eukaryotic chromosome
Repeat unit
– T T A G G G T T A G G G T T A G G G T T A G G G 3′
– A A T C C C A A T
C C C A A U C C C
RNA
3′
5′
Telomerase synthesizes a 6-nucleotide repeat.
Telomerase
G
G
– T T A G G G T T A G G G T T A G G G T T A G G G T T A G
– A A T C C C A A T
C C C A A U C C C
Telomerase moves 6 nucleotides to the right and begins to make another repeat.
T
T
– T T A G G G T T A G G G T T A G G G T T A G G G T T A G G G
– A A T C C C A A T
C C C A A U C C C
The complementary strand is made by primase, DNA polymerase, and ligase.
– T T A G G G T T A G G G T T A G G G T T A G G G T T A G G G T T A G G G 3′
– A A T C C C A A T C C C A A T C C C A A T C C C A A U C C C A A U
RNA primer

ONLINE ANIMATION

FIGURE 13.22 **The enzymatic action of telomerase.** A short, three-nucleotide segment of RNA within telomerase causes it to bind to the 3′ overhang. The adjacent part of the RNA is used as a template to make a short, six-nucleotide repeat of DNA. After the repeat is made, telomerase moves six nucleotides to the right and then synthesizes another repeat. This process is repeated many times to lengthen the top strand shown in this figure. The bottom strand is made by DNA polymerase, using an RNA primer at the end of the chromosome that is complementary to the telomeric repeat sequence in the top strand. DNA polymerase fills in the region, which is sealed by ligase.

Concept check: *How many times would telomerase have to bind to a different site in the telomere to make a segment of DNA that is 36 nucleotides in length?*

- DNA polymerase is unable to replicate the very end of a eukaryotic chromosome (see Figure 13.21).
- Telomerase uses a short RNA molecule as a template to add repeat sequences onto telomeres (see Figure 13.22).

13.5 COMPREHENSION QUESTIONS

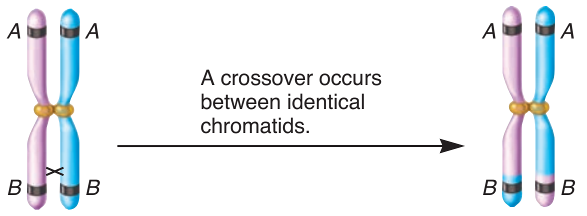
1. In eukaryotes, DNA replication is initiated at an origin of replication by
   a. DnaA.
   b. the origin recognition complex.
   c. DNA polymerase δ.
   d. MCM helicase.
2. Which of the following statements regarding DNA polymerases in eukaryotes is *not* correct?
   a. DNA polymerase α synthesizes a short RNA–DNA primer.
   b. DNA polymerases δ and ε synthesize most of the lagging and leading strands, respectively.
   c. Lesion-replicating DNA polymerases can replicate over damaged DNA.
   d. All of the above statements are correct.
3. In eukaryotes, RNA primers are primarily removed by
   a. DNA polymerase I.
   b. DNA polymerase α.
   c. flap endonuclease.
   d. helicase.
4. To synthesize DNA, what does telomerase use as a template?
   a. It uses the DNA in the 3′ overhang.
   b. It uses RNA that is already a component of telomerase.
   c. No template is used.
   d. Both a and b are correct.

13.6 HOMOLOGOUS RECOMBINATION

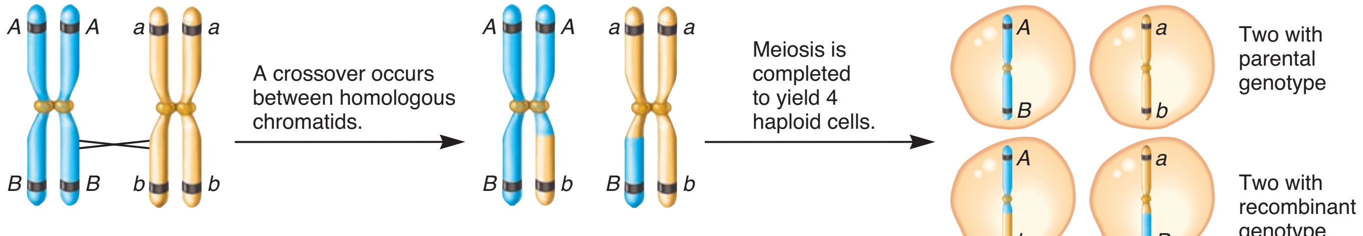
Homologous recombination involves an exchange between DNA segments that are similar or identical in their DNA sequences. Eukaryotic chromosomes that have similar or identical sequences frequently participate in crossing over during meiosis I and occasionally during mitosis. Crossing over involves the alignment of a pair of homologous chromosomes, followed by the breakage of two chromatids at analogous locations, and the subsequent exchange of the corresponding segments (refer to Figure 2.10).

Figure 13.23 shows two types of crossing over that may occur between replicated chromosomes in a diploid species. When crossing over takes place between sister chromatids, the process is called **sister chromatid exchange (SCE).** Because sister chromatids are genetically identical, SCE does not produce a new combination of alleles (Figure 13.23a). By comparison, it is common for crossing over to occur between homologous chromosomes during meiosis. As shown in Figure 13.23b, this form of homologous recombination may produce a new combination of alleles in the resulting chromosomes. In this second case, homologous recombination has resulted in **genetic recombination,** which refers to the shuffling of genetic material to create a new genetic combination that differs from the original. Homologous recombination is an important mechanism that fosters genetic variation.

Bacteria are usually haploid. They do not have pairs of homologous chromosomes. Even so, bacteria also can undergo homologous recombination. How can the exchange of DNA segments occur in a haploid organism? First, bacteria may have more than one copy of a chromosome per cell, though the copies are usually identical. These copies can exchange genetic material via homologous recombination. Second, during DNA replication,



(a) Sister chromatid exchange


A
A
a
a
A crossover occurs between homologous chromatids.
B
B
b
b
A
A
a
a
B
b
B
b
Meiosis is completed to yield 4 haploid cells.
A
B
a
b
Two with parental genotype
A
b
a
B
Two with recombinant genotype

(b) Recombination between homologous chromosomes

FIGURE 13.23 **Crossing over between eukaryotic chromosomes.** **(a)** Sister chromatid exchange occurs when genetically identical chromatids cross over. **(b)** Homologous recombination can also occur when homologous chromatids cross over. This form of homologous recombination may lead to a new combination of alleles, which is called a recombinant (or nonparental) genotype.

Genes→Traits Homologous recombination is particularly important when we consider the relationships between multiple genes and multiple traits. For example, if the X chromosome in a female fruit fly carried alleles for red eyes and gray body, and its homolog carried alleles for white eyes and yellow body, homologous recombination could produce recombinant chromosomes that carry alleles for red eyes and yellow body, or alleles for white eyes and gray body. Therefore, new combinations of two or more alleles can arise when homologous recombination takes place.

Concept check: *What is the advantage of genetic recombination, which is depicted in part b?*

the replicated regions may also undergo homologous recombination. In bacteria, homologous recombination is particularly important in the repair of DNA segments that have been damaged. In this section, we will focus our attention on the molecular mechanisms that underlie homologous recombination.

The Holliday Model Describes a Molecular Mechanism for the Recombination Process

It is perhaps surprising that the first molecular model of homologous recombination did not come from a biochemical analysis of DNA or from electron microscopy studies. Instead, it was deduced from the outcome of genetic crosses in fungi.

Geneticists have learned a great deal from the analysis of fungi, because certain species produce asci (singular, ascus), which are sacs that contain the products of a single meiosis. When two haploid fungi that differ at a single gene are crossed to each other, the ascus is expected to contain an equal proportion of each genotype. For example, if a pigmented strain of *Neurospora* producing orange spores is crossed to an albino strain producing white spores, the resulting group of eight cells, or octad, should contain four orange spores and four white spores. In 1934, H. Zickler noticed that unequal proportions of the spores sometimes occurred within asci. He occasionally observed octads with six orange spores and two white spores, or six white spores and two orange spores.

Zickler used the term **gene conversion** to describe the phenomenon in which one allele is converted to the allele on the homologous chromosome. Subsequent studies by several researchers confirmed this phenomenon in yeast and *Neurospora*. Gene conversion occurred at too high a rate to be explained by new mutations. In addition, research showed that gene conversion often occurs in a chromosomal region where a crossover has taken place.

Based on studies involving gene conversion, Robin Holliday proposed a model in 1964 to explain the molecular steps that occur during homologous recombination. We will first consider the steps in the Holliday model and then consider more recent models. Later, we will examine how the Holliday model can explain the phenomenon of gene conversion.

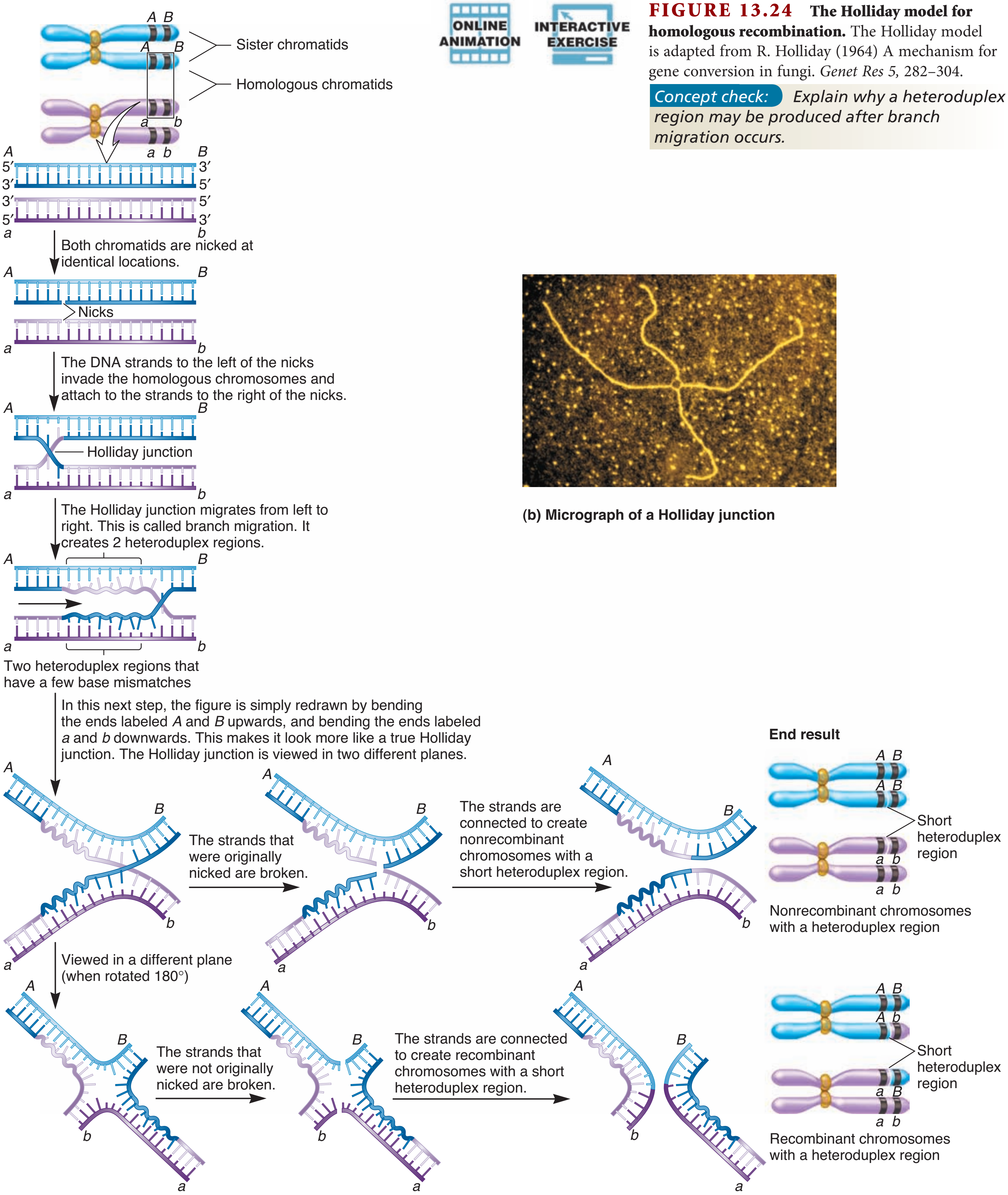
The **Holliday model** is shown in **Figure 13.24a**. At the beginning of the process, two homologous chromatids are aligned with each other. According to the model, a break or nick occurs at identical sites in one strand of each of the two homologous chromatids. The strands then invade the opposite helices and base-pair with the complementary strands. This event is followed by a covalent linkage to create a **Holliday junction.** The cross in the Holliday junction can migrate in a lateral direction. As it does so, a DNA strand in one helix is swapped for a DNA strand in the other helix. This process is called **branch migration** because the branch connecting the two double helices migrates laterally. Because the DNA sequences in the homologous chromosomes are similar but may not be identical, the swapping of the DNA strands during branch migration may produce a **heteroduplex,** a region in the double-stranded DNA that contains base mismatches. In other words, because the DNA strands in this region are from homologous chromosomes, their sequences may not be perfectly complementary and therefore yield mismatches.

ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 13.24 The Holliday model for homologous recombination. The Holliday model is adapted from R. Holliday (1964) A mechanism for gene conversion in fungi. *Genet Res* 5, 282–304.

Concept check: *Explain why a heteroduplex region may be produced after branch migration occurs.*


A B
A B
Sister chromatids
Homologous chromatids
a b
a b
A
B
5′
3′
3′
5′
3′
5′
5′
3′
a
b
Both chromatids are nicked at identical locations.
A
B
Nicks
a
b
The DNA strands to the left of the nicks invade the homologous chromosomes and attach to the strands to the right of the nicks.
A
B
Holliday junction
a
b
The Holliday junction migrates from left to right. This is called branch migration. It creates 2 heteroduplex regions.
A
B
a
b
Two heteroduplex regions that have a few base mismatches
In this next step, the figure is simply redrawn by bending the ends labeled A and B upwards, and bending the ends labeled a and b downwards. This makes it look more like a true Holliday junction. The Holliday junction is viewed in two different planes.
A
B
a
b
The strands that were originally nicked are broken.
A
B
a
b
The strands are connected to create nonrecombinant chromosomes with a short heteroduplex region.
A
B
a
b
End result
A B
A B
Short heteroduplex region
a b
a b
Nonrecombinant chromosomes with a heteroduplex region
Viewed in a different plane (when rotated 180°)
A
B
b
a
The strands that were not originally nicked are broken.
A
B
b
a
The strands are connected to create recombinant chromosomes with a short heteroduplex region.
A
B
b
a
A B
A b
Short heteroduplex region
a B
a b
Recombinant chromosomes with a heteroduplex region

(b) Micrograph of a Holliday junction

(a) The Holliday model for homologous recombination

The final two steps in the recombination process are collectively called **resolution** because they involve the breakage and rejoining of two DNA strands to produce two separate chromosomes. In other words, the entangled DNA strands become resolved into two separate structures. The bottom left side of Figure 13.24a shows the Holliday junction viewed from two different planes. If breakage occurs in the same two DNA strands that were originally nicked at the beginning of this process, the subsequent joining of strands produces nonrecombinant chromosomes with a heteroduplex region. Alternatively, if breakage occurs in the strands that were not originally nicked, the rejoining process results in recombinant chromosomes, also with a heteroduplex region.

The Holliday model can account for the general properties of recombinant chromosomes formed during eukaryotic meiosis. As mentioned, the original model was based on the results of crosses in fungi in which the products of meiosis are contained within a single ascus. Nevertheless, molecular research in many other organisms has supported the central tenets of the model. Particularly convincing evidence came from electron microscopy studies in which recombination structures could be visualized. **Figure 13.24b** shows an electron micrograph of two DNA fragments in the process of recombination. This structure has been called a chi form because its shape is similar to the Greek letter χ (chi).

More Recent Models Have Refined the Molecular Steps of Homologous Recombination

As more detailed studies of homologous recombination have become available, certain steps in the Holliday model have been reconsidered. In particular, more recent models have modified the initiation phase of recombination. Researchers now suggest that homologous recombination is not likely to involve nicks at identical sites in one strand of each homologous chromatid. Instead, it is more likely for a DNA helix to incur a break in both strands of one chromatid or a single nick. Both of these types of changes have been shown to initiate homologous recombination. Therefore, newer models have tried to incorporate these experimental observations. In 1975, a model proposed by Matthew Meselson and Charles Radding hypothesized that a single nick in one DNA strand initiates recombination. A second model, proposed by Jack Szostak, Terry Orr-Weaver, Rodney Rothstein, and Franklin Stahl, suggests that a double-strand break initiates the recombination process. This is called the **double-strand break model.** Though recombination may occur via more than one mechanism, research suggests that double-strand breaks commonly promote homologous recombination during meiosis and during DNA repair.

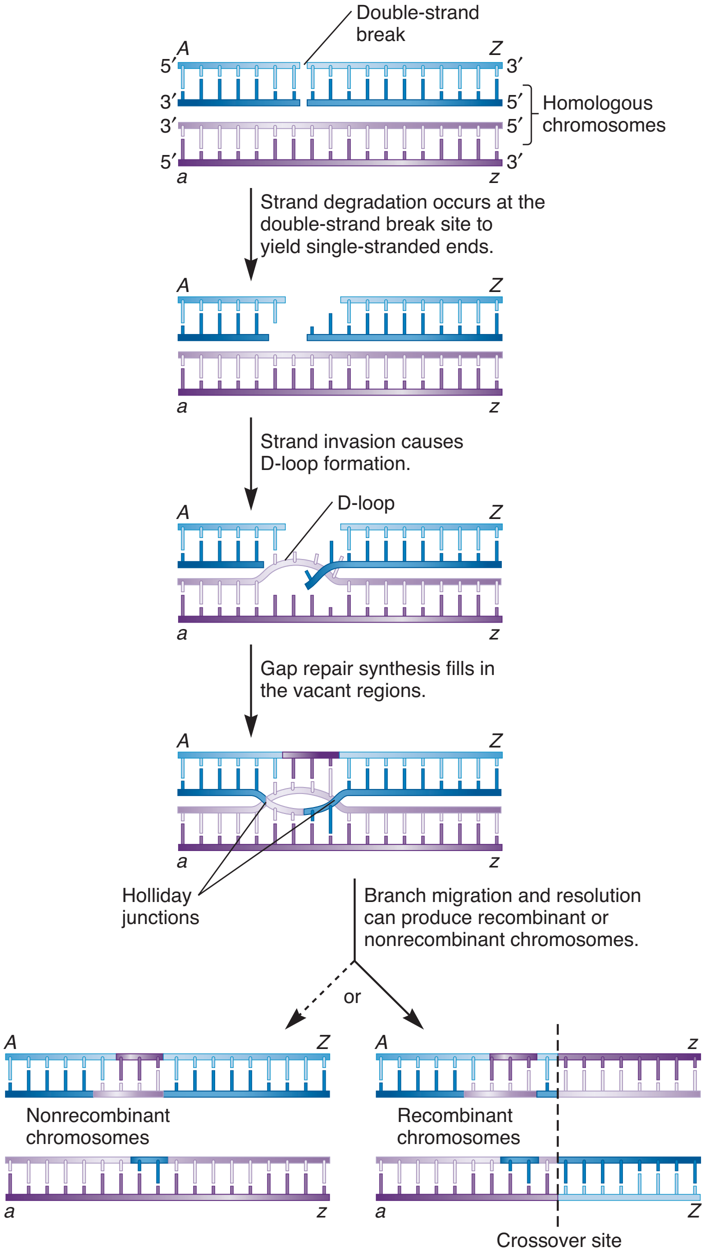

Double-strand break
A
Z
5′
3′
3′
5′
3′
5′
5′
3′
Homologous chromosomes
a
z
Strand degradation occurs at the double-strand break site to yield single-stranded ends.
A
Z
a
z
Strand invasion causes D-loop formation.
D-loop
A
Z
a
z
Gap repair synthesis fills in the vacant regions.
A
Z
a
z
Holliday junctions
Branch migration and resolution can produce recombinant or nonrecombinant chromosomes.
or
A
Z
Nonrecombinant chromosomes
a
z
A
z
Recombinant chromosomes
a
Z
Crossover site

FIGURE 13.25 **A simplified version of the double-strand break model.** For simplicity, this illustration does not include the formation of heteroduplexes. The dashed arrow indicates that the pathway to the left may be less favored.

Concept check: Describe the structure and location of a D-loop.

Figure 13.25 shows the general steps in the double-strand break model. As seen here, the top chromosome has experienced a double-strand break. A small region near the break is degraded, which generates a single-stranded DNA segment that can invade the intact double helix. The strand displaced by the invading segment forms a structure called a displacement loop (D-loop). After the D-loop is formed, two regions have a gap in the DNA. How is the problem fixed? DNA synthesis occurs in the relatively short gaps where a DNA strand is missing. This DNA synthesis is called **DNA gap repair synthesis.** Once this is completed, two Holliday junctions are produced. Depending on the way these are resolved, the end result is nonrecombinant or recombinant chromosomes containing a short heteroduplex. In eukaryotes such as yeast, evidence suggests that certain proteins bound to Holliday junctions may regulate the resolution step in a way that favors the formation of recombinant chromosomes rather than nonrecombinant chromosomes.

Various Proteins Are Necessary to Facilitate Homologous Recombination

The homologous recombination process requires the participation of many proteins that catalyze different steps in the recombination pathway. Homologous recombination is found in all species, and the types of proteins that participate in the steps outlined in Figure 13.25 are very similar. The cells of any given species may have more than one molecular mechanism to carry out homologous recombination. This process is best understood in *Escherichia coli*. **Table 13.5** summarizes some of the *E. coli* proteins that play critical roles in one recombination pathway found in this species. Though it is beyond the scope of this textbook, *E. coli* has other pathways to carry out homologous recombination.

RecBCD is a protein complex composed of the RecB, RecC, and RecD proteins. (The term *Rec* indicates that these proteins are involved with recombination.) The RecBCD complex plays an important role in the initiation of recombination involving double-strand breaks, as outlined in Figure 13.25. In this process, RecBCD recognizes a double-strand break within DNA and catalyzes DNA unwinding and strand degradation. The action of RecBCD produces single-stranded DNA ends that can participate in strand invasion and exchange. The single-stranded DNA ends are coated with single-stranded binding protein to prevent their further degradation. The RecBCD complex can also create breaks in the DNA at sites known as chi sequences. In *E. coli*, the chi sequence is 5′–GCTGGTGG–3′. As RecBCD tracks along the DNA, when it encounters a chi sequence, it cuts one of the DNA strands to the 3′ side of this sequence. This generates a single nick that can also initiate homologous recombination.

The function of the RecA protein is to promote strand invasion. To accomplish this task, it binds to the single-stranded ends of DNA molecules generated from the activity of RecBCD. A large number of RecA proteins bind to single-stranded DNA, forming a structure called a filament. During strand invasion, this filament makes contact with the unbroken chromosome. Initially, this contact is most likely to occur at nonhomologous regions. The contact point slides along the DNA until it reaches a homologous region. Once a homologous site is located, RecA catalyzes the displacement of one DNA strand, and the invading single-stranded DNA quickly forms a double helix with the other strand. This results in a D-loop, as shown in Figure 13.25. RecA proteins mediate the movement of the invading strand and the displacement of the complementary strand. This occurs in such a way that the displaced strand invades the vacant region of the broken chromosome.

Proteins that bind specifically to Holliday junctions have also been identified. These include a complex of proteins termed RuvABC and a protein called RecG. RuvA is a tetramer that forms a platform on which the Holliday junction is held in a square planar configuration. This platform also contains two hexameric rings of RuvB. Together, RuvA and RuvB act as a helicase that catalyzes an ATP-dependent migration of a four-way branched DNA junction in either the 5′ → 3′ or 3′ → 5′ direction. RuvC is an endonuclease that binds as a dimer to Holliday junctions and resolves these structures by making cuts in the DNA (see bottom of Figure 13.24a). RecG is a helicase that also catalyzes branch migration of Holliday junctions.

TABLE **13.5**

***E. coli* Proteins That Play a Role in Homologous Recombination**

Protein	Description
RecBCD	A complex of three proteins that tracks along the DNA and recognizes double-strand breaks. The complex partially degrades the double-stranded regions to generate single-stranded regions that can participate in strand invasion. RecBCD is also involved in loading RecA onto single-stranded DNA. In addition, RecBCD can create single-strand breaks at chi sequences.
Single-stranded binding protein	Coats broken ends of chromosomes and prevents excessive strand degradation.
RecA	Binds to single-stranded DNA and promotes strand invasion, which enables homologous strands to find each other. It also promotes the displacement of the complementary strand to generate a D-loop.
RuvABC	This protein complex binds to Holliday junctions. RuvAB promotes branch migration. RuvC is an endonuclease that cuts the crossed or uncrossed strands to resolve Holliday junctions into separate chromosomes.
RecG	RecG protein can also promote branch migration of Holliday junctions.

Gene Conversion May Result from DNA Mismatch Repair or DNA Gap Repair

As mentioned earlier, homologous recombination can lead to gene conversion, in which one allele is converted to the allele on the homologous chromosome. The original Holliday model was based on this phenomenon.

How can homologous recombination account for gene conversion? Researchers have identified two possible ways this can occur. One mechanism involves DNA mismatch repair, a topic described in Chapter 18. To understand how this works, let's take a closer look at the heteroduplexes formed during branch migration of a Holliday junction (see Figure 13.24a). A heteroduplex contains a DNA strand from each of the two original parental chromosomes. The two parental chromosomes may contain an allelic difference within this region. In other words, this short region may contain DNA sequence differences. If this is the case, the heteroduplex formed after branch migration will contain an area of base mismatch. Gene conversion occurs when recombinant chromosomes are repaired and result in two copies of the same allele.

As shown in **Figure 13.26**, mismatch repair of a heteroduplex may result in gene conversion. In this example, the two parental chromosomes had different alleles due to a single base-pair difference in their DNA sequences, as shown at the top of the figure. During recombination, branch migration has occurred across this region, thereby creating two heteroduplexes with base

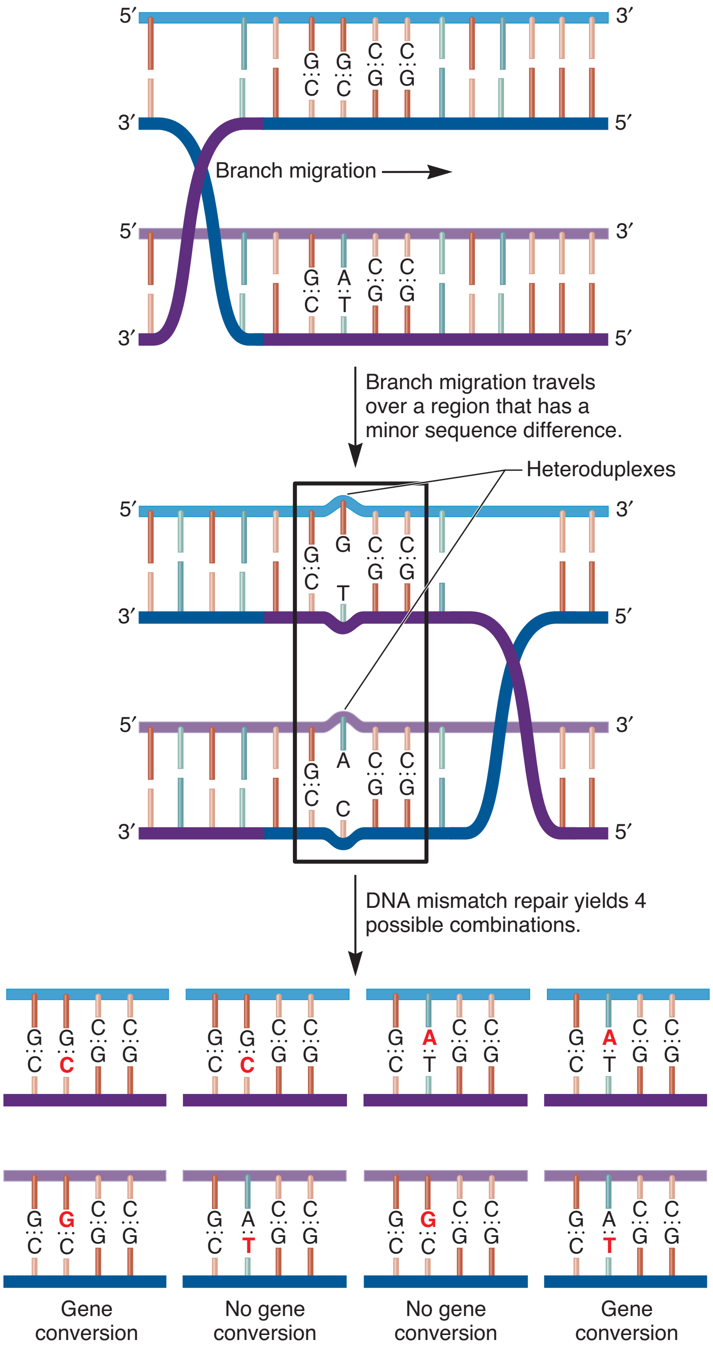

5′
3′
3′
5′
G G C C
C C G G
Branch migration
5′
3′
G A C C
C T G G
3′
5′
Branch migration travels over a region that has a minor sequence difference.
Heteroduplexes
DNA mismatch repair yields 4 possible combinations.
Gene conversion
No gene conversion
No gene conversion
Gene conversion

FIGURE 13.26 **Gene conversion by DNA mismatch repair.** A branch migrates past a homologous region that contains slightly different DNA sequences. This produces two heteroduplexes: DNA double helices with mismatches. The mismatches can be repaired in four possible ways by the mismatch repair system described in Chapter 18. Two of these ways result in gene conversion. The repaired base is shown in red.

mismatches. As described in Chapter 18, DNA mismatches are recognized by DNA repair systems and repaired to a double helix that obeys the AT/GC rule. These two mismatches can be repaired in four possible ways. As shown here, two possibilities produce no gene conversion, whereas the other two lead to gene conversion.

A second mechanism for gene conversion occurs via DNA gap repair synthesis. Figure 13.27 illustrates how gap repair synthesis can lead to gene conversion according to the double-strand

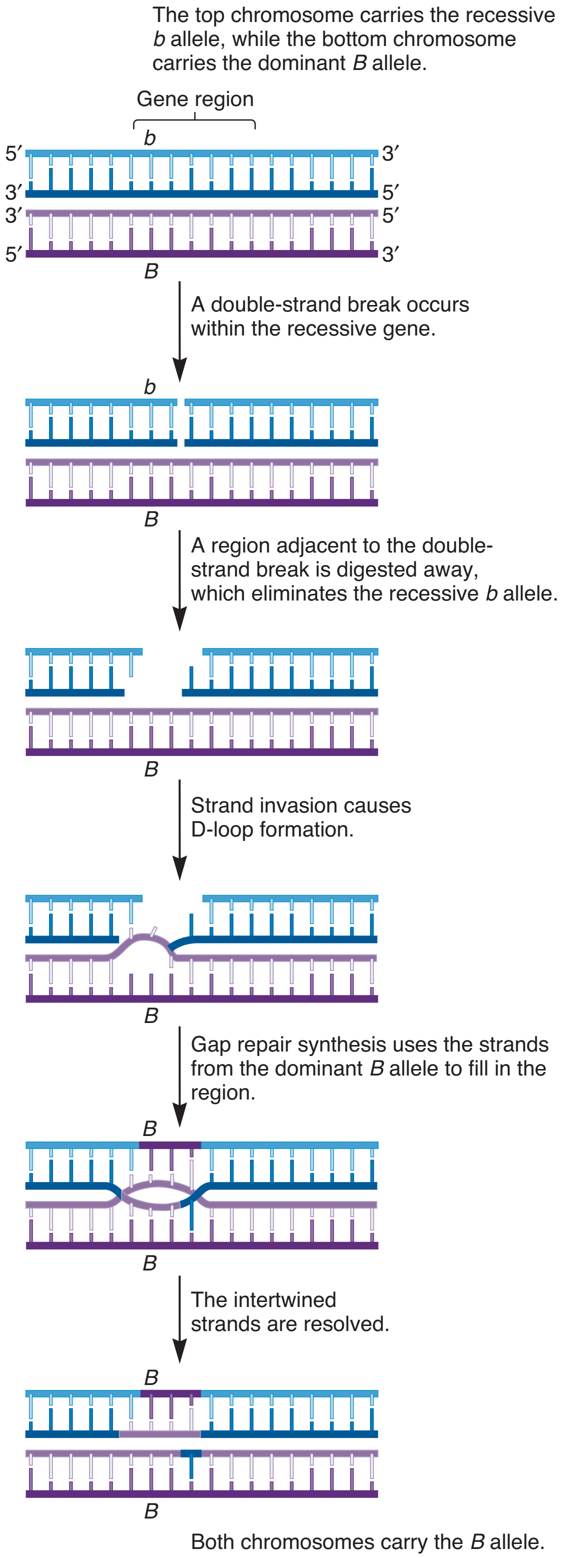

The top chromosome carries the recessive *b* allele, while the bottom chromosome carries the dominant *B* allele.
Gene region
b
5′
3′
3′
5′
3′
5′
5′
3′
B
A double-strand break occurs within the recessive gene.
b
B
A region adjacent to the double-strand break is digested away, which eliminates the recessive *b* allele.
B
Strand invasion causes D-loop formation.
B
Gap repair synthesis uses the strands from the dominant *B* allele to fill in the region.
B
B
The intertwined strands are resolved.
B
B
Both chromosomes carry the *B* allele.

FIGURE 13.27 **Gene conversion by gap repair synthesis in the double-strand break model.** A gene is found in two alleles, designated *B* and *b*. A double-strand break occurs in the DNA encoding the *b* allele. Both of these DNA strands are digested away, thereby eliminating the *b* allele. A complementary DNA strand encoding the *B* allele migrates to this region and provides the template to synthesize a double-stranded region. Following resolution, both DNA double helices carry the *B* allele.

Concept check: Explain what happened to the b allele that allowed gene conversion to occur.

break model. The top chromosome, which carries the recessive *b* allele, has suffered a double-strand break in this gene. A gap is created by the digestion of the DNA in the double helix. This digestion eliminates the *b* allele. The two template strands used in gap repair synthesis are from one homologous chromatid. This helix carries the dominant *B* allele. Therefore, after gap repair synthesis takes place, the top chromosome carries the *B* allele, as does the bottom chromosome. Gene conversion has changed the recessive *b* allele to a dominant *B* allele.

13.6 REVIEWING THE KEY CONCEPTS

- Homologous recombination involves an exchange between DNA segments that are similar or identical in their DNA sequences. It can occur between sister chromatids or between homologous chromosomes (see Figure 13.23).
- The Holliday model describes the molecular steps that occur during homologous recombination between homologous chromosomes (see Figure 13.24).
- The initiation of homologous recombination is thought to usually occur via a double-strand break (see Figure 13.25).
- Several different proteins are involved with homologous recombination (see Table 13.5).
- Two different mechanisms, DNA mismatch repair and gap repair synthesis, can explain how gene conversion may occur during homologous recombination (see Figures 13.26, 13.27).

13.6 COMPREHENSION QUESTIONS

1. Homologous recombination refers to the exchange between DNA segments that are
   a. similar or identical in their DNA sequences.
   b. in close proximity to one another.
   c. broken due to ionizing radiation.
   d. misaligned along a chromosome.
2. During the molecular process of homologous recombination between homologous chromosomes,
   a. a Holliday junction forms.
   b. branch migration occurs.
   c. a heteroduplex region forms.
   d. all of the above occur.
3. A key difference between the original Holliday model and the double-strand break model is the way that
   a. the DNA strands are initially broken.
   b. branch migration occurs.
   c. a heteroduplex is formed
   d. resolution occurs.
4. Which of the following mechanisms can cause gene conversion?
   a. DNA mismatch repair
   b. DNA gap repair
   c. The resolution of a Holliday junction
   d. Both a and b

KEY TERMS

Page 268. DNA replication, homologous recombination
Page 269. template strands, parental strands
Page 270. daughter strands, conservative model, semiconservative model, dispersive model
Page 273. origin of replication, bidirectionally, replication fork
Page 274. DnaA proteins, DnaA box sequences, DNA helicase
Page 275. bidirectional replication
Page 276. topoisomerase II, DNA gyrase, single-strand binding proteins, RNA primers, primase, leading strand, lagging strand, DNA polymerase
Page 278. Okazaki fragments, DNA ligase, primosome, replisome, dimeric DNA polymerase
Page 279. termination sequences
Page 280. catenanes
Page 282. processive enzyme, fidelity
Page 283. proofreading, ARS elements
Page 284. prereplication complex (preRC), origin recognition complex (ORC), MCM helicase, DNA replication licensing, lesion-replicating polymerases
Page 285. flap endonuclease, telomeres
Page 286. telomerase, telomerase reverse transcriptase (TERT)
Page 287. sister chromatid exchange (SCE), genetic recombination
Page 288. gene conversion, Holliday model, Holliday junction, branch migration, heteroduplex
Page 290. resolution, double-strand break model, DNA gap repair synthesis

CHAPTER SUMMARY

- DNA replication is the process in which existing DNA strands are used to make new DNA strands.

13.1 Structural Overview of DNA Replication

- DNA replication occurs when the strands of DNA unwind and each strand is used as a template to make a new strand according to the AT/GC rule. The resulting DNA molecules have the same base sequence as the original DNA (see Figure 13.1).
- By labeling DNA with heavy and light isotopes of nitrogen and using centrifugation, Meselson and Stahl showed that DNA replication is semiconservative (see Figures 13.2, 13.3).

13.2 Bacterial DNA Replication: The Formation of Two Replication Forks at the Origin of Replication

- Bacterial DNA replication begins at a single origin of replication and proceeds bidirectionally around the circular chromosome (see Figure 13.4).
- In *E. coli*, DNA replication is initiated when DnaA proteins bind to five DnaA boxes at the origin of replication and cause the AT-rich region to unwind. DNA helicases continue to sepa-

rate the DNA strands, thereby promoting the movement of two replication forks (see Figures 13.5, 13.6).

13.3 Bacterial DNA Replication: The Synthesis of New DNA Strands

- At each replication fork, DNA helicase unwinds the DNA and topoisomerase II alleviates positive supercoiling. Single-strand binding proteins coat the DNA to prevent the strands from coming back together. Primase synthesizes RNA primers and DNA polymerase synthesizes complementary strands of DNA. DNA ligase seals the gaps between Okazaki fragments (see Figure 13.7, Table 13.1).
- DNA polymerase III in *E. coli* is an enzyme with several subunits. The catalytic subunit wraps around the DNA like a hand (see Figure 13.8, Table 13.2).
- DNA polymerase needs a primer to synthesize DNA and makes new DNA strands in a 5′ to 3′ direction (see Figure 13.9).
- During DNA synthesis, the leading strand is made continuously in the direction of the replication fork, whereas the lagging strand is made as Okazaki fragments in the direction away from the fork (see Figures 13.10, 13.11).
- The primosome is a complex between helicase and primase. The replisome is a complex between the primosome and dimeric DNA polymerase (see Figure 13.12).
- In *E. coli*, DNA replication is terminated at ter sequences (see Figure 13.13).
- Following DNA replication, interlocked catenanes sometimes need to be unlocked via topoisomerase II (see Figure 13.14).

13.4 Bacterial DNA Replication: Chemistry and Accuracy

- DNA polymerase III is a processive enzyme that uses deoxyribonucleoside triphosphates (dNTPs) to make new DNA strands (see Figure 13.15).
- The high fidelity of DNA replication is a result of (1) the stability of hydrogen bonding between the correct bases, (2) the phenomenon of induced fit, and (3) the proofreading ability of DNA polymerase (see Figure 13.16).

13.5 Eukaryotic DNA Replication

- Eukaryotic chromosomes contain multiple origins of replication. Part of the prereplication complex is formed from a group of six proteins called the origin recognition complex (ORC). The binding of MCM helicase completes a process called DNA replication licensing (see Figures 13.17, 13.18).
- Eukaryotes have several different DNA polymerases with specialized roles. Different types of DNA polymerases switch with each other during the process of DNA replication (see Table 13.3).
- Flap endonuclease is an enzyme that removes RNA primers from Okazaki fragments (see Figure 13.19).
- The ends of eukaryotic chromosomes contain telomeres, which are composed of short repeat sequences and proteins (see Figure 13.20, Table 13.4).
- DNA polymerase is unable to replicate the very end of a eukaryotic chromosome (see Figure 13.21).
- Telomerase uses a short RNA molecule as a template to add repeat sequences onto telomeres (see Figure 13.22).

13.6 Homologous Recombination

- Homologous recombination involves an exchange between DNA segments that are similar or identical in their DNA sequences. It can occur between sister chromatids or between homologous chromosomes (see Figure 13.23).
- The Holliday model describes the molecular steps that occur during homologous recombination between homologous chromosomes (see Figure 13.24).
- The initiation of homologous recombination is thought to usually occur via a double-strand break (see Figure 13.25).
- Several different proteins are involved with homologous recombination (see Table 13.5).
- Two different mechanisms, DNA mismatch repair and gap repair synthesis, can explain how gene conversion may occur during homologous recombination (see Figures 13.26, 13.27).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe three ways to account for the high fidelity of DNA replication. Discuss the quantitative contributions of each of the three ways.

Answer:

First: AT and GC pairs are preferred in the double-helix structure. This provides fidelity to around 1 mistake per 1000.

Second: Induced fit by DNA polymerase prevents covalent bond formation unless the proper nucleotides are in place. This increases fidelity another 100- to 1000-fold, to about 1 error in 100,000 to 1 million.

Third: Exonuclease proofreading increases fidelity another 100- to 1000-fold, to about 1 error per 100 million nucleotides added.

S2. What do you think would happen if the ter sequences were deleted from the bacterial DNA?

Answer: Instead of meeting at the ter sequences, the two replication forks would meet somewhere else. This would depend on how fast they were moving. For example, if one fork was advancing faster than the other, they would meet closer to where the slower-moving fork started. In fact, researchers have actually conducted this experiment. Interestingly, *E. coli* without the ter sequences seemed to survive just fine.

S3. Summarize the steps that occur in the process of chromosomal DNA replication in *E. coli*.

Answer:

Step 1. DnaA proteins bind to the origin of replication, resulting in the separation of the AT-rich region.

Step 2. DNA helicase breaks the hydrogen bonds between the DNA strands, topoisomerase II alleviates positive supercoiling, and single-strand binding proteins hold the parental strands apart.

Step 3. Primase synthesizes one RNA primer in the leading strand and many RNA primers in the lagging strand. DNA polymerase III then synthesizes the daughter strands of DNA. In the lagging strand, many short segments of DNA (Okazaki fragments) are made. DNA polymerase I removes the RNA primers and fills in with DNA, and DNA ligase covalently links the Okazaki fragments together.

Step 4. The processes described in steps 2 and 3 continue until the two replication forks reach the ter sequences on the other side of the circular bacterial chromosome.

Step 5. Topoisomerases unravel the intertwined chromosomes if necessary.

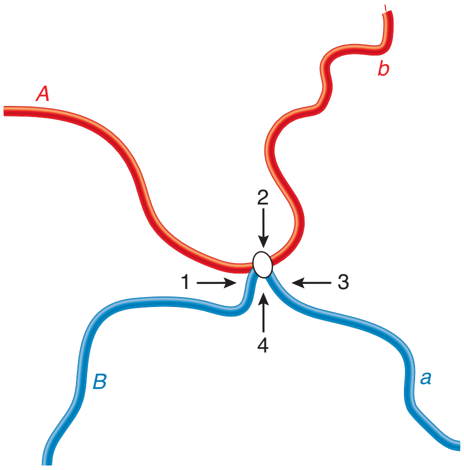
S4. Zickler was the first person to demonstrate gene conversion by observing unusual ratios in *Neurospora* octads. At first, it was difficult for geneticists to believe these results because they seemed to contradict the Mendelian concept that alleles do not physically interact with each other. However, work by Mary Mitchell provided convincing evidence that gene conversion actually takes place. She investigated three different genes in *Neurospora*. One *Neurospora* strain had three mutant alleles: *pdx-1* (pyridoxine-requiring), *pyr-1* (pyrimidine-requiring), and *col-4* (a mutation that affected growth morphology). The *pdx-l* gene had been previously shown to map between the *pyr-1* and *col-4* genes. Mitchell crossed this strain to a wild-type *Neurospora* strain:

pyr-1 pdx-1 col-4 × *pyr-1*$^+$ *pdx-1*$^+$ *col-4*$^+$

She first analyzed many octads with regard to their requirement for pyridoxine. Out of 246 octads, 2 had an aberrant ratio in which two spores were *pdx-1,* and six were *pdx-1*$^+$. These same spores were then analyzed with regard to the other two genes. In both cases, the aberrant asci gave a normal 4:4 ratio of *pyr-1:pyr-1*$^+$ and *col-4:col-4*$^+$. Explain these results.

Answer: These results can be explained by gene conversion. The gene conversion took place in a limited region of the chromosome (within the *pdx-1* gene), but it did not affect the flanking genes (*pyr-1* and *col-4*) located on either side of the *pdx-1* gene. In the asci containing two *pdx-1* alleles and six *pdx-1*$^+$ alleles, a crossover occurred during meiosis I in the region of the *pdx-1* gene. Gene conversion changed the *pdx-1* allele into the *pdx-1*$^+$ allele. This gene conversion could have occurred by two mechanisms. If branch migration occurred across the *pdx-1* gene, a heteroduplex may have formed, which could be repaired by DNA mismatch repair, as described in Figure 13.26. In the aberrant asci with two *pdx-1* and six *pdx-1*$^+$ alleles, the *pdx-1* allele was converted to *pdx-1*$^+$. Alternatively, gene conversion of *pdx-1* into *pdx-1*$^+$ could have taken place via gap repair synthesis, as described in Figure 13.27. In this case, the *pdx-1* allele would have been digested away, and the DNA encoding the *pdx-1*$^+$ allele would have migrated into the digested region and provided a template to make a copy of the *pdx-1*$^+$ allele. (*Note:* Since this pioneering work, additional studies have shed considerable light on the phenomenon of gene conversion. It occurs at a fairly high rate in fungi, approximately 0.1 to 1% of the time. It is not due to new mutations occurring during meiosis.)

S5. A schematic drawing of an uncrossed Holliday junction is shown here. One chromatid is shown in red, and the homologous chromatid is shown in blue. The red chromatid carries a dominant allele labeled *A* and a recessive allele labeled *b*, whereas the blue chromatid carries a recessive allele labeled *a* and a dominant allele labeled *B*.



Where would the breakage of the crossed strands have to occur to get recombinant chromosomes? Would it have to occur at sites 1 and 3, or at sites 2 and 4? What would be the genotypes of the two recombinant chromosomes?

Answer:

Breakage would have to occur at the arrows labeled 2 and 4. This would connect the *A* allele with the *B* allele. The *a* allele in the other homolog would become connected with the *b* allele. In other words, one chromosome would be *AB,* and the homolog would be *ab*.

Conceptual Questions

C1. What key structural features of the DNA molecule underlie its ability to be faithfully replicated?

C2. With regard to DNA replication, define the term *bidirectional replication.*

C3. Which of the following statements is not true? Explain why.

A. A DNA strand can serve as a template strand on many occasions.

B. Following semiconservative DNA replication, one strand is a newly made daughter strand and the other strand is a parental strand.

C. A DNA double helix may contain two strands of DNA that were made at the same time.

D. A DNA double helix obeys the AT/GC rule.

E. A DNA double helix could contain one strand that is 10 generations older than its complementary strand.

C4. The compound known as nitrous acid is a reactive chemical that replaces amino groups (–NH_2) with keto groups (=O). When nitrous acid reacts with the bases in DNA, it can change cytosine to uracil and change adenine to hypoxanthine. A DNA double helix has the following sequence:

TTGGATGCTGG
AACCTACGACC

A. What would be the sequence of this double helix immediately after reaction with nitrous acid? Let the letter H represent hypoxanthine and U represent uracil.

B. Let's suppose this DNA was reacted with nitrous acid. The nitrous acid was then removed, and the DNA was replicated for two generations. What would be the sequences of the DNA products after the DNA had replicated twice? Your answer should contain the sequences of four double helices. *Note:* During DNA replication, uracil hydrogen bonds with adenine, and hypoxanthine hydrogen bonds with cytosine.

C5. One way that bacterial cells regulate DNA replication is by GATC methylation sites within the origin of replication. Would this mechanism work if the DNA was conservatively (rather than semiconservatively) replicated?

C6. The chromosome of *E. coli* contains 4.6 million bp. How long will it take to replicate its DNA? Assuming DNA polymerase III is the primary enzyme involved and this enzyme can actively proofread during DNA synthesis, how many base pair mistakes will be made in one round of DNA replication in a bacterial population containing 1000 bacteria?

C7. Here are two strands of DNA.

—————————————————DNA polymerase–>

—————————————————————————————————

The one on the bottom is a template strand, and the one on the top is being synthesized by DNA polymerase in the direction shown by the arrow. Label the 5′ and 3′ ends of the top and bottom strands.

C8. A DNA strand has the following sequence:

5′–GATCCCGATCCGCATACATTTACCAGATCACCACC–3′

In which direction would DNA polymerase slide along this strand (from left to right or from right to left)? If this strand was used as a template by DNA polymerase, what would be the sequence of the newly made strand? Indicate the 5′ and 3′ ends of the newly made strand.

C9. List and briefly describe the three types of sequences within bacterial origins of replication that are functionally important.

C10. As shown in Figure 13.5, five DnaA boxes are found within the origin of replication in *E. coli*. Take a look at these five sequences carefully.

A. Are the sequences of the five DnaA boxes very similar to each other? (*Hint:* Remember that DNA is double-stranded; think about these sequences in the forward and reverse direction.)

B. What is the most common sequence for the DnaA box? In other words, what is the most common base in the first position, second position, and so on until the ninth position? The most common sequence is called the consensus sequence.

C. The *E. coli* chromosome is about 4.6 million bp long. Based on random chance, is it likely that the consensus sequence for a DnaA box occurs elsewhere in the *E. coli* chromosome? If so, why aren't there multiple origins of replication in *E. coli*?

C11. Obtain two strings of different colors (e.g., black and white) that are the same length. A length of 20 inches is sufficient. Tie a knot at one end of the black string, and tie a knot at one end of the white string. Each knot designates the 5′ end of your strings. Make a double helix with your two strings. Now tape one end of the double helix to a table so that the tape is covering the knot on the black string.

A. Pretend your hand is DNA helicase and use your hand to unravel the double helix, beginning at the end that is not taped to the table. Should your hand be sliding along the white string or the black string?

B. As in Figure 13.13, imagine that your two hands together form dimeric DNA polymerase. Unravel your two strings halfway to create a replication fork. Grasp the black string with your left hand and the white string with your right hand. Your thumbs should point toward the 5′ end of each string. You need to loop one of the strings so that one of the DNA polymerases can synthesize the lagging strand. With such a loop, the dimeric DNA polymerase can move toward the replication fork and synthesize both DNA strands in the 5′ to 3′ direction. In other words, with such a loop, your two hands can touch each other with both of your thumbs pointing toward the fork. Should the black string be looped, or should the white string be looped?

C12. Sometimes DNA polymerase makes a mistake, and the wrong nucleotide is added to the growing DNA strand. With regard to pyrimidines and purines, two general types of mistakes are possible. The addition of an incorrect pyrimidine instead of the correct pyrimidine (e.g., adding cytosine where thymine should be added) is called a transition. If a pyrimidine is incorrectly added to the growing strand instead of purine (e.g., adding cytosine where an adenine should be added), this type of mistake is called a transversion. If a transition or transversion is not detected by DNA polymerase, a mutation is created that permanently changes the DNA sequence. Though both types of mutations are rare, transition mutations are more frequent than transversion mutations. Based on your understanding of DNA replication and DNA polymerase, offer three explanations why transition mutations are more common.

C13. A short genetic sequence, which may be recognized by DNA primase, is repeated many times throughout the *E. coli* chromosome. Researchers have hypothesized that DNA primase may recognize this sequence as a site to begin the synthesis of an RNA primer for DNA replication. The *E. coli* chromosome is roughly 4.6 million bp in length. How many copies of the DNA primase recognition sequence would be necessary to replicate the entire *E. coli* chromosome?

C14. Single-strand binding proteins keep the two parental strands of DNA separated from each other until DNA polymerase has an opportunity to replicate the strands. Suggest how single-strand binding proteins keep the strands separated and yet do not impede the ability of DNA polymerase to replicate the strands.

C15. The ability of DNA polymerase to digest a DNA strand from one end is called its exonuclease activity. Exonuclease activity is used to digest RNA primers and also to proofread a newly made DNA strand. *Note:* DNA polymerase I does not change direction while it is removing an RNA primer and synthesizing new DNA. It does change direction during proofreading.

A. In which direction, 5′ to 3′ or 3′ to 5′ is the exonuclease activity occurring during the removal of RNA primers and during the proofreading and removal of mistakes following DNA replication?

B. Figure 13.16 shows a drawing of the 3′ exonuclease site. Do you think this site would be used by DNA polymerase I to remove RNA primers? Why or why not?

C16. In the following drawing (see next page), the top strand is the template DNA, and the bottom strand shows the lagging strand prior to the action of DNA polymerase I. The lagging strand contains three Okazaki fragments. The RNA primers, which are designated as a series of asterisks, have not yet been removed.

The top strand is the template DNA.

3′—————————————————————————————————5′
5′*************____***************____*************_____3′
RNA primer ↑ RNA primer ↑ RNA primer

|———————————||———————————||———————————|
Left Okazaki fragment Middle Okazaki fragment Right Okazaki fragment

A. Which Okazaki fragment was made first, the one on the left or the one on the right?

B. Which RNA primer would be the first one to be removed by DNA polymerase I, the primer on the left or the primer on the right? For this primer to be removed by DNA polymerase I and for the gap to be filled in, is it necessary for the Okazaki fragment in the middle to have already been synthesized? Explain why.

C. Let's consider how DNA ligase connects the left Okazaki fragment with the middle Okazaki fragment. After DNA polymerase I removes the middle RNA primer and fills in the gap with DNA, where does DNA ligase function? See the arrows on either side of the middle RNA primer. Is ligase needed at the left arrow, at the right arrow, or both?

D. When connecting two Okazaki fragments, DNA ligase needs to use NAD^+ or ATP as a source of energy to catalyze this reaction. Explain why DNA ligase needs another source of energy to connect two nucleotides, but DNA polymerase needs nothing more than the incoming nucleotide and the existing DNA strand. *Note:* You may want to refer to Figure 13.15 to answer this question.

C17. Describe the three important functions of the DnaA protein.

C18. Draw a picture that illustrates how DNA helicase works.

C19. What is an Okazaki fragment? In which strand of DNA are Okazaki fragments found? Based on the properties of DNA polymerase, why is it necessary to make these fragments?

C20. Discuss the similarities and differences in the synthesis of DNA in the lagging and leading strands. What is the advantage of a primosome and a replisome as opposed to having all replication enzymes functioning independently of each other?

C21. Explain the proofreading function of DNA polymerase.

C22. What is a processive enzyme? Explain why this is an important feature of DNA polymerase.

C23. What enzymatic features of DNA polymerase prevent it from replicating one of the DNA strands at the ends of linear chromosomes? Compared with DNA polymerase, how is telomerase different in its ability to synthesize a DNA strand? What does telomerase use as its template for the synthesis of a DNA strand? How does the use of this template result in a telomere sequence that is tandemly repetitive?

C24. As shown in Figure 13.21, telomerase attaches additional DNA, six nucleotides at a time, to the ends of eukaryotic chromosomes. However, it works in only one DNA strand. Describe how the opposite strand is replicated.

C25. If a eukaryotic chromosome has 25 origins of replication, how many replication forks does it have at the beginning of DNA replication?

C26. A diagram of a linear chromosome is shown here (see upper right column). The end of each strand is labeled with an A, B, C, or D. Which ends could not be replicated by DNA polymerase? Why not?

5′–A————————————————————————————B–3′
3′–C————————————————————————————D–5′

C27. As discussed in Chapter 10, some viruses contain RNA as their genetic material. Certain RNA viruses can exist as a provirus in which the viral genetic material has been inserted into the chromosomal DNA of the host cell. For this to happen, the viral RNA must be copied into a strand of DNA. An enzyme called reverse transcriptase, encoded by the viral genome, copies the viral RNA into a complementary strand of DNA. The strand of DNA is then used as a template to make a double-stranded DNA molecule. This double-stranded DNA molecule is then inserted into the chromosomal DNA, where it may exist as a provirus for a long period of time.

A. How is the function of reverse transcriptase similar to the function of telomerase?

B. Unlike DNA polymerase, reverse transcriptase does not have a proofreading function. How might this affect the proliferation of the virus?

C28. Telomeres contain a 3′ overhang region, as shown in Figure 13.20. Does telomerase require a 3′ overhang to replicate the telomere region? Explain.

C29. Describe the similarities and differences between homologous recombination involving sister chromatid exchange (SCE) and that involving homologs. Would you expect the same types of proteins to be involved in both processes? Explain.

C30. The molecular mechanism of SCE is similar to homologous recombination between homologs except that the two segments of DNA are sister chromatids instead of homologous chromatids. If branch migration occurs during SCE, will a heteroduplex be formed? Explain why or why not. Can gene conversion occur during sister chromatid exchange?

C31. Which steps in the double-strand break model for recombination would be inhibited if the following proteins were missing? Explain the function of each protein required for the step that is inhibited.

A. RecBCD

B. RecA

C. RecG

D. RuvABC

C32. What two molecular mechanisms can explain the phenomenon of gene conversion? Would both occur in the double-strand break model?

C33. Is homologous recombination an example of mutation? Explain.

C34. What are recombinant chromosomes? How do they differ from the original parental chromosomes from which they are derived?

C35. In the Holliday model for homologous recombination (see Figure 13.24), the resolution steps can produce recombinant or nonrecombinant chromosomes. Explain how this can occur.

C36. What is gene conversion?

C37. Make a list of the differences between the Holliday model and the double-strand break model.

C38. In recombinant chromosomes, where is gene conversion likely to take place: near the breakpoint or far away from the breakpoint? Explain.

C39. What events does the RecA protein facilitate?

C40. According to the double-strand break model, does gene conversion necessarily involve DNA mismatch repair? Explain.

C41. What type of DNA structure is recognized by RecG and RuvABC? Do you think these proteins recognize DNA sequences? Be specific about what type(s) of molecular recognition these proteins can perform.

Application and Experimental Questions

E1. Answer the following questions that pertain to the experiment of Figure 13.3.

A. What would be the expected results if the Meselson and Stahl experiment was carried out for four or five generations?

B. What would be the expected results of the Meselson and Stahl experiment after three generations if the mechanism of DNA replication was dispersive?

C. As shown in the data, explain why three different bands (i.e., light, half-heavy, and heavy) can be observed in the CsCl gradient.

E2. An absent-minded researcher follows the steps of Figure 13.3, and when the gradient is viewed under UV light, the researcher does not see any bands at all. Which of the following mistakes could account for this observation? Explain how.

A. The researcher forgot to add ^{14}N-containing compounds.

B. The researcher forgot to add lysozyme.

C. The researcher forgot to turn on the UV lamp.

E3. Figure 13.4b shows an autoradiograph of a replicating bacterial chromosome. If you analyzed many replicating chromosomes, what types of information could you learn about the mechanism of DNA replication?

E4. The technique of dideoxy sequencing of DNA is described in Chapter 19. The technique relies on the use of dideoxyribonucleotides (shown in Figures 19.16 and 19.17). A dideoxyribonucleotide has a hydrogen atom attached to the 3′-carbon atom instead of an –OH group. When a dideoxyribonucleotide is incorporated into a newly made strand, the strand cannot grow any longer. Explain why.

E5. Another technique described in Chapter 19 is polymerase chain reaction (PCR) (see Figure 19.4). This method is based on our understanding of DNA replication. In this method, a small amount of double-stranded template DNA is mixed with a high concentration of primers. Nucleotides and DNA polymerase are also added. The template DNA strands are separated by heat treatment, and when the temperature is lowered, the primers can bind to the single-stranded DNA, and then DNA polymerase replicates the DNA. This increases the amount of DNA made from the primers. This cycle of steps (i.e., heat treatment, lower temperature, allow DNA replication to occur) is repeated over and over. Because the cycles are repeated many times, this method is called a chain reaction. It is called a polymerase chain reaction because DNA polymerase is the enzyme needed to increase the amount of DNA with each cycle. In a PCR experiment, the template DNA is placed in a tube, and the primers, nucleotides, and DNA polymerase are added to the tube. The tube is then placed in a machine called a thermocycler, which raises and lowers the temperature. During one cycle, the temperature is raised (e.g., to 95°C) for a brief period and then lowered (e.g., to 60°C) to allow the primers to bind. The sample is incubated at a slightly higher temperature for a few minutes to allow DNA replication to proceed. In a typical PCR experiment, the tube may be left in the thermocycler for 25 to 30 cycles. The total time for a PCR experiment is a few hours.

A. Why is DNA helicase not needed in a PCR experiment?

B. How is the sequence of each primer important in a PCR experiment? Do the two primers recognize the same strand or opposite strands?

C. The DNA polymerase used in PCR experiments is a DNA polymerase isolated from thermophilic bacteria. Why is this kind of polymerase used?

D. If a tube initially contained 10 copies of double-stranded DNA, how many copies of double-stranded DNA (in the region flanked by the two primers) would be obtained after 27 cycles?

Questions for Student Discussion/Collaboration

1. The complementarity of double-stranded DNA is the underlying reason that DNA can be faithfully copied. Propose alternative chemical structures that could be faithfully copied.
2. DNA replication is fast, virtually error-free, and coordinated with cell division. Discuss which of these three features you think is the most important.
3. If no homologous recombination of any kind could occur, what would be the harmful and beneficial consequences?

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

Answers to Comprehension Questions

13.1: c, d, b

13.2: b, d

13.3: a, d

13.4: a, d

13.5: b, d, c, b

13.6: a, d, a, d

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

PART IV MOLECULAR PROPERTIES OF GENES

A molecular model showing the enzyme RNA polymerase (gray and yellow) in the act of sliding along the DNA (blue) and synthesizing a copy of RNA (red).

14

CHAPTER OUTLINE

14.1 Overview of Transcription

14.2 Transcription in Bacteria

14.3 Transcription in Eukaryotes

14.4 RNA Modification

GENE TRANSCRIPTION AND RNA MODIFICATION

The primary function of the genetic material, which is DNA, is to store the information necessary to create a living organism. The information is contained within units called genes. At the molecular level, a **gene** is defined as a segment of DNA that is used to make a functional product—either an RNA molecule or a polypeptide. How is the information within a gene accessed? The first step in this process is called **transcription,** which literally means the act or process of making a copy. In genetics, this term refers to the process of synthesizing RNA from a DNA sequence (**Figure 14.1**). The structure of DNA is not altered as a result of transcription. Rather, the DNA base sequence has only been accessed to make a copy in the form of RNA. Therefore, the same DNA can continue to store information. DNA replication, which was discussed in Chapter 13, provides a mechanism for copying that information so it can be transmitted from cell to cell and from parent to offspring.

Structural genes encode the amino acid sequence of a polypeptide. When a structural gene is transcribed, the first product is an RNA molecule known as **messenger RNA (mRNA).** During polypeptide synthesis—a process called **translation**—the sequence of nucleotides within the mRNA determines the sequence of amino acids in a polypeptide. One or more polypeptides then assemble into a functional protein. The synthesis of functional proteins ultimately determines an organism's traits. The model depicted in Figure 14.1, which is called the **central dogma of genetics** (also called the central dogma of molecular biology), was first enunciated by Francis Crick in 1958. The central dogma of genetics specifies that the usual flow of genetic information is from DNA to mRNA to polypeptide. It forms a cornerstone of our understanding of genetics at the molecular level.

In this chapter, we begin to study the molecular steps in **gene expression,** the overall process by which the information within a gene is used to produce a functional product such as a polypeptide. Our emphasis will be on transcription and the modifications that may occur to an RNA transcript after it has been made. In Chapter 15, we will examine the process of translation, and Chapters 16 and 17 will focus on how gene expression is regulated at the molecular level.

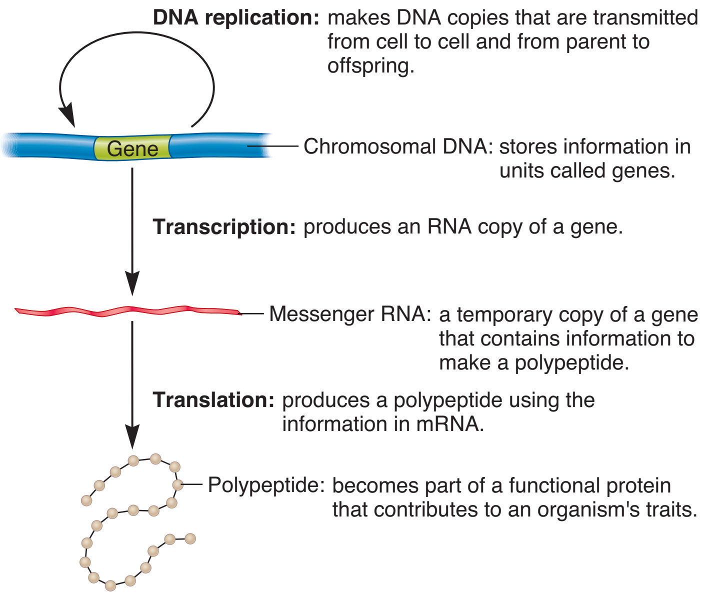

DNA replication: makes DNA copies that are transmitted from cell to cell and from parent to offspring.
Gene
Chromosomal DNA: stores information in units called genes.
Transcription: produces an RNA copy of a gene.
Messenger RNA: a temporary copy of a gene that contains information to make a polypeptide.
Translation: produces a polypeptide using the information in mRNA.
Polypeptide: becomes part of a functional protein that contributes to an organism's traits.

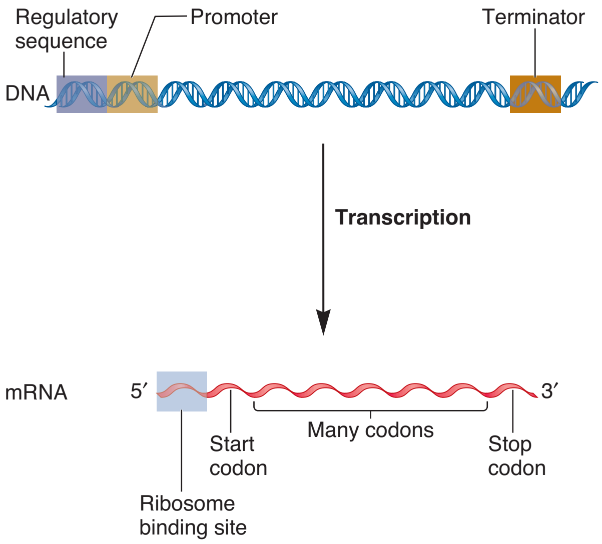
FIGURE 14.1 **The central dogma of genetics.** The usual flow of genetic information is from DNA to mRNA to polypeptide. Note: The direction of informational flow shown in this figure is the most common direction, but exceptions occur. For example, RNA viruses and certain transposable elements use an enzyme called reverse transcriptase to make a copy of DNA from RNA.

14.1 OVERVIEW OF TRANSCRIPTION

One key concept that is important in the process of transcription is that short base sequences define the beginning and ending of a gene and also play a role in regulating the level of RNA synthesis. In this section, we begin by examining the sequences that determine where transcription starts and ends; we also briefly consider DNA sequences, called regulatory sites, that influence whether a gene is turned on or off. A second important concept is the role of proteins in transcription. DNA sequences, in and of themselves, just exist. For genes to be actively transcribed, proteins must recognize particular DNA sequences and act on them in a way that affects the transcription process. In the later part of this section, we will consider how proteins participate in the general steps of transcription and the types of RNA transcripts that can be made.

Gene Expression Requires Base Sequences That Perform Different Functional Roles

The molecular expression of genes, along with environmental factors, determines an organism's traits. For a gene to be expressed, a few different types of base sequences perform specific roles. Figure 14.2 shows a common organization of base sequences needed to create a structural gene that functions in a bacterium such as *E. coli*. Each type of base sequence performs its role during a particular stage of gene expression. For example, the promoter and terminator are base sequences used


Regulatory sequence
Promoter
Terminator
DNA
Transcription
mRNA
5′
3′
Start codon
Many codons
Stop codon
Ribosome binding site

DNA:

- **Regulatory sequences:** site for the binding of regulatory proteins; the role of regulatory proteins is to influence the rate of transcription. Regulatory sequences can be found in a variety of locations.
- **Promoter:** site for RNA polymerase binding; signals the beginning of transcription.
- **Terminator:** signals the end of transcription.

mRNA:

- **Ribosome-binding site:** site for ribosome binding; translation begins near this site in the mRNA. In eukaryotes, the ribosome scans the mRNA for a start codon.
- **Start codon:** specifies the first amino acid in a polypeptide sequence, usually a formylmethionine (in bacteria) or a methionine (in eukaryotes).
- **Codons:** 3-nucleotide sequences within the mRNA that specify particular amino acids. The sequence of codons within mRNA determines the sequence of amino acids within a polypeptide.
- **Stop codon:** specifies the end of polypeptide synthesis.
- A bacterial mRNA may be polycistronic, which means it encodes two or more polypeptides.

FIGURE 14.2 **Organization of sequences of a bacterial gene and its mRNA transcript.** This figure depicts the general organization of sequences that are needed to create a functional gene that encodes an mRNA.

Concept check: *If a mutation changed the start codon into a stop codon, would this mutation affect the length of the RNA? Explain.*

during gene transcription. Specifically, the **promoter** provides a site to begin transcription, and the **terminator** specifies the end of transcription. These two sequences cause RNA synthesis to occur within a defined location. As shown in Figure 14.2, the DNA is transcribed into RNA from the end of the promoter to the terminator.

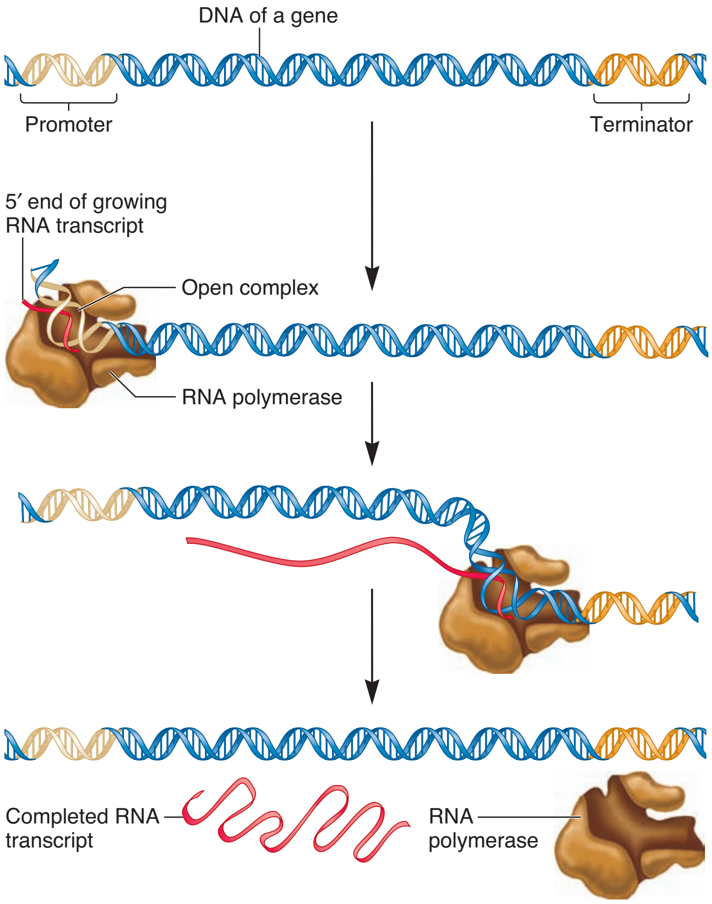
Proteins called **transcription factors** recognize base sequences in the DNA and control transcription. Some transcription factors bind directly to the promoter and facilitate transcription. Other transcription factors recognize **regulatory sequences,** or **regulatory elements**—short stretches of DNA involved in the regulation of transcription. Certain transcription factors bind to such regulatory sequences and increase the rate of transcription, whereas others inhibit transcription.

Base sequences within an mRNA are used during the translation process. In bacteria, a short sequence within the mRNA, the **ribosome-binding site,** provides a location for the ribosome to bind and begin translation. The bacterial ribosome recognizes this site because it is complementary to a sequence in ribosomal RNA. In addition, mRNA contains a series of **codons,** read as groups of three nucleotides, which contain the information for a polypeptide's sequence. The first codon, which is very close to the ribosome-binding site, is the **start codon.** This is followed by many more codons that dictate the sequence of amino acids within the synthesized polypeptide. Finally, a **stop codon** signals the end of translation.

The Three Stages of Transcription Are Initiation, Elongation, and Termination

Transcription occurs in three stages: **initiation; elongation,** or synthesis of the RNA transcript; and **termination** (Figure 14.3). These steps involve protein–DNA interactions in which proteins such as **RNA polymerase,** the enzyme that synthesizes RNA, interacts with DNA sequences. What causes transcription to begin? The initiation stage in the transcription process is a recognition step. The sequence of bases within the promoter region is recognized by transcription factors. The specific binding of transcription factors to the promoter sequence identifies the starting site for transcription.

Transcription factors and RNA polymerase first bind to the promoter region when the DNA is in the form of a double helix.



Initiation: The promoter functions as a recognition site for transcription factors (not shown). The transcription factor(s) enables RNA polymerase to bind to the promoter. Following binding, the DNA is denatured into a bubble known as the open complex.

Elongation/synthesis of the RNA transcript: RNA polymerase slides along the DNA in an open complex to synthesize RNA.

Termination: A terminator is reached that causes RNA polymerase and the RNA transcript to dissociate from the DNA.

ONLINE ANIMATION

FIGURE 14.3 Stages of transcription.

Genes→Traits The ability of genes to produce an organism's traits relies on the molecular process of gene expression. Transcription is the first step in gene expression. During transcription, the gene's sequence within the DNA is used as a template to make a complementary copy of RNA. In Chapter 15, we examine how the sequence in mRNA is translated into a polypeptide chain. After polypeptides are made within a living cell, they fold into functional proteins that govern an organism's traits.

For transcription to occur, the DNA strands must be separated. This allows one of the two strands to be used as a template for the synthesis of a complementary strand of RNA. This synthesis occurs as RNA polymerase slides along the DNA, forming a small bubble-like structure known as the **open complex.** Eventually, RNA polymerase reaches a terminator, which causes both RNA polymerase and the newly made RNA transcript to dissociate from the DNA.

RNA Transcripts Have Different Functions

Once they are made, RNA transcripts play different functional roles (**Table 14.1**). Well over 90% of all genes are structural genes, which are transcribed into mRNA. For structural genes, mRNAs are made first, but the final, functional products are polypeptides that are components of proteins. The remaining types of RNAs described in Table 14.1 are never translated. The RNA transcripts from such nonstructural genes have various important cellular functions. For nonstructural genes, the functional product is the RNA. In some cases, the RNA transcript becomes part of a complex that contains both protein subunits and one or more RNA molecules. Examples of protein–RNA complexes include ribosomes, signal recognition particles, RNaseP, spliceosomes, and telomerase.

14.1 REVIEWING THE KEY CONCEPTS

- A gene is an organization of DNA sequences. A promoter signals the start of transcription, and a terminator signals the end. Regulatory sequences control the rate of transcription. For genes that encode polypeptides, the gene sequence also specifies a start codon, a stop codon, and many codons in between. Bacterial genes also specify a ribosome-binding sequence (see Figure 14.2).
- Transcription occurs in three phases called initiation, elongation, and termination (see Figure 14.3).
- RNA transcripts have many different functions (see Table 14.1).

14.1 COMPREHENSION QUESTIONS

1. Which of the following base sequences would be used during transcription?
   a. Promoter and terminator
   b. Start and stop codons
   c. Ribosome-binding sequence
   d. Both a and b
2. The three stages of transcription are
   a. initiation, ribosome-binding, and termination.
   b. elongation, ribosome-binding, and termination.
   c. initiation, elongation, and termination.
   d. initiation, regulation, termination.
3. Which of the following statements is *false*?
   a. All genes encode polypeptides that function as units within proteins.
   b. Some genes encode RNAs that are never translated.
   c. Most genes encode mRNAs.
   d. Some RNAs function within protein–RNA complexes.

TABLE **14.1**

Functions of RNA Molecules

Type of RNA	Description
mRNA	Messenger RNA (mRNA) encodes the sequence of amino acids within a polypeptide. In bacteria, some mRNAs encode a single polypeptide. Other mRNAs are polycistronic—a single mRNA encodes two or more polypeptides. In most species of eukaryotes, each mRNA usually encodes a single polypeptide. However, in some species, such as *Caenorhabditis elegans* (a nematode worm), polycistronic mRNAs are relatively common.
tRNA	Transfer RNA (tRNA) is necessary for the translation of mRNA. The structure and function of tRNA are outlined in Chapter 15.
rRNA	Ribosomal RNA (rRNA) is necessary for the translation of mRNA. Ribosomes are composed of both rRNAs and protein subunits. The structure and function of ribosomes are examined in Chapter 15.
MicroRNA	MicroRNAs (miRNAs) are short RNA molecules that are involved in gene regulation in eukaryotes (see Chapter 17).
scRNA	Small cytoplasmic RNA (scRNA) is found in the cytoplasm of bacteria and eukaryotes. In bacteria, scRNA is needed for protein secretion. An example in eukaryotes is 7S RNA, which is necessary in the targeting of proteins to the endoplasmic reticulum. It is a component of a complex known as signal recognition particle (SRP), which is composed of 7S RNA and six different protein subunits.
RNA of RNaseP	RNaseP is necessary in the processing of all bacterial tRNA molecules. The RNA is the catalytic component of RNaseP, which is composed of a 350- to 410-nucleotide RNA and one protein subunit.
snRNA	Small nuclear RNA (snRNA) is necessary in the splicing of eukaryotic pre-mRNA. snRNAs and proteins combine to form small nuclear riboproteins (snRNPs), which are components of a spliceosome. The structure and function of spliceosomes are examined later in this chapter.
Telomerase RNA	The enzyme telomerase, which is involved in the replication of eukaryotic telomeres, is composed of an RNA molecule and protein subunits (see Chapter 13).
snoRNA	Small nucleolar RNA (snoRNA) is necessary in the processing of eukaryotic rRNA transcripts. snoRNAs are also associated with protein subunits. In eukaryotes, snoRNAs are found in the nucleolus, where rRNA processing and ribosome assembly occur.
Viral RNAs	Some types of viruses use RNA as their genome, which is packaged within the viral capsid.

14.2 TRANSCRIPTION IN BACTERIA

Our molecular understanding of gene transcription initially came from studies involving bacteria and bacteriophages. Several early investigations focused on the production of viral RNA after bacteriophage infection. The first suggestion that RNA is derived from the transcription of DNA was made by Elliot Volkin and Lazarus Astrachan in 1956. When the researchers exposed *E. coli* cells to T2 bacteriophage, they observed that the RNA made immediately after infection had a base composition substantially different from the base composition of RNA prior to infection. Furthermore, the base composition after infection was very similar to the base composition in the T2 DNA, except that the RNA contained uracil instead of thymine. These results were consistent with the idea that the bacteriophage DNA is used as a template for the synthesis of bacteriophage RNA.

In 1960, Matthew Meselson and François Jacob found that proteins are synthesized on ribosomes. One year later, Jacob and his colleague Jacques Monod proposed that a certain type of RNA acts as a genetic messenger (from the DNA to the ribosome) to provide the information for protein synthesis. They hypothesized that this RNA, which they called messenger RNA (mRNA), is transcribed from the sequence within DNA and then directs the synthesis of particular polypeptides. In the early 1960s, this proposal was remarkable, considering that it was made before the actual isolation and characterization of the mRNA molecules in vitro. In 1961, the hypothesis was confirmed by Sydney Brenner in collaboration with Jacob and Meselson. They found that when a virus infects a bacterial cell, a virus-specific RNA is made that rapidly associates with preexisting ribosomes in the cell.

Since these pioneering studies, a great deal has been learned about the molecular features of bacterial gene transcription. Much of our knowledge comes from studies in *E. coli*. In this section, we will examine the three steps in the gene transcription process as they occur in bacteria.

A Promoter Is a Short Sequence of DNA That Is Necessary to Initiate Transcription

The type of DNA sequence known as the promoter gets its name from the idea that it "promotes" gene expression. More precisely, this sequence of bases directs the exact location for the initiation of RNA transcription. Most of the promoter region is located just ahead of or upstream from the site where transcription of a gene actually begins. By convention, the bases in a promoter sequence are numbered in relation to the **transcriptional start site** (Figure 14.4). This site is the first base used as a template for RNA transcription and is denoted +1. The bases preceding this site are numbered in a negative direction. No base is numbered zero. Therefore, most of the promoter region is labeled with negative numbers that describe the number of bases preceding the beginning of transcription.

Although the promoter may encompass a region several dozen nucleotides in length, short **sequence elements** are particularly critical for promoter recognition. By comparing the sequence of DNA bases within many promoters, researchers have learned that certain sequences of bases are necessary to create a functional promoter. In many promoters found in *E. coli* and similar species, two sequence elements are important. These are located at approximately the −35 and −10 sites in the promoter region (see Figure 14.4). The sequence in the top DNA strand at the −35 region is 5′–TTGACA–3′, and the one at the −10 region is 5′–TATAAT–3′. The TATAAT sequence is sometimes called the **Pribnow box** after David Pribnow, who initially discovered it in 1975.

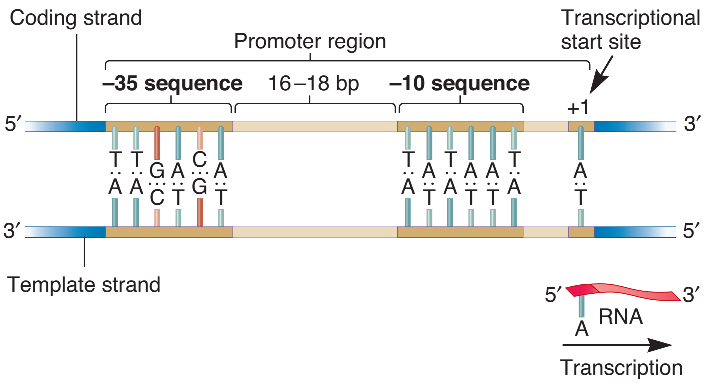
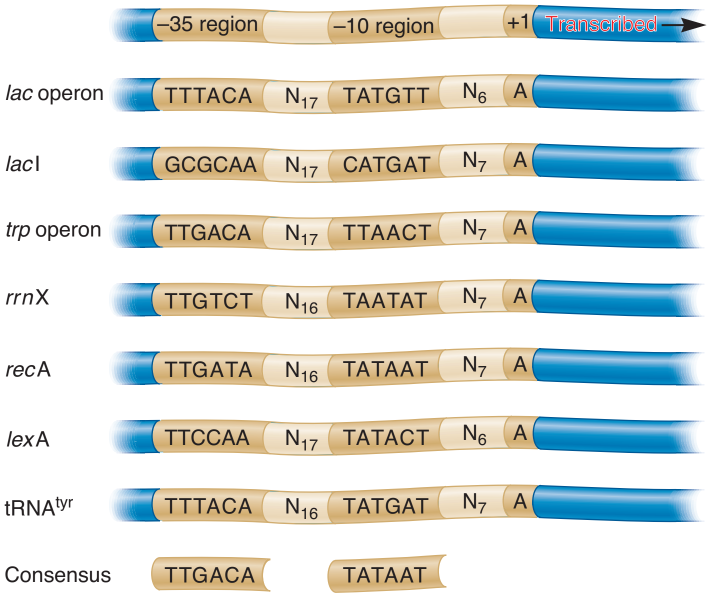


FIGURE 14.4 **The conventional numbering system of promoters.** The first nucleotide that acts as a template for transcription is designated +1. The numbering of nucleotides to the left of this spot is in a negative direction, whereas the numbering to the right is in a positive direction. For example, the nucleotide that is immediately to the left of the +1 nucleotide is numbered –1, and the nucleotide to the right of the +1 nucleotide is numbered +2. There is no zero nucleotide in this numbering system. In many bacterial promoters, sequence elements at the –35 and –10 regions play a key role in promoting transcription.



lac operon	TTTACA	N_{17}	TATGTT	N_6	A
*lac*I	GCGCAA	N_{17}	CATGAT	N_7	A
trp operon	TTGACA	N_{17}	TTAACT	N_7	A
*rrn*X	TTGTCT	N_{16}	TAATAT	N_7	A
*rec*A	TTGATA	N_{16}	TATAAT	N_7	A
*lex*A	TTCCAA	N_{17}	TATACT	N_6	A
$tRNA^{tyr}$	TTTACA	N_{16}	TATGAT	N_7	A
Consensus	TTGACA		TATAAT		

FIGURE 14.5 **Examples of –35 and –10 sequences within a variety of bacterial promoters.** This figure shows the –35 and –10 sequences for one DNA strand found in seven bacterial and bacteriophage promoters. The consensus sequence is at the bottom. The spacer regions contain the designated number of nucleotides between the –35 and –10 region or between the –10 region and the transcriptional start site. For example, N_{17} means there are 17 nucleotides between the end of the –35 region and the beginning of the –10 region.

Concept check: *What does the term consensus sequence mean?*

The sequences at the −35 and −10 sites can vary among different genes. For example, Figure 14.5 illustrates the sequences found in several different *E. coli* promoters. The most commonly occurring bases within a sequence element form the **consensus sequence.** This sequence is efficiently recognized by proteins that initiate transcription. For many bacterial genes, a strong correlation is found between the maximal rate of RNA transcription and the degree to which the −35 and −10 regions agree with their consensus sequences. Mutations in the −35 or −10 region that alter consensus sequences typically slow down the rate of transcription.

Bacterial Transcription Is Initiated When RNA Polymerase Holoenzyme Binds at a Promoter Sequence

Thus far, we have considered the DNA sequences that constitute a functional promoter. Let's now turn our attention to the proteins that recognize those sequences and carry out the transcription process. The enzyme that catalyzes the synthesis of RNA is RNA polymerase. In *E. coli*, the **core enzyme** is composed of five subunits, $\alpha_2\beta\beta'\varpi$. The association of a sixth subunit, **sigma** (σ) **factor,** with the core enzyme is referred to as RNA polymerase **holoenzyme.** The different subunits within the holoenzyme play distinct functional roles. The two α subunits are important in the proper assembly of the holoenzyme and in the process of binding to DNA. The β and β′ subunits are also needed for binding to the DNA and carry out the catalytic synthesis of RNA. The ϖ (omega) subunit is important for the proper assembly of the core enzyme. The holoenzyme is required to initiate transcription; the primary role of σ factor is to recognize the promoter. Proteins, such as σ factor, that influence the function of RNA polymerase are types of transcription factors.

After RNA polymerase holoenzyme is assembled into its six subunits, it binds loosely to the DNA and then slides along the DNA, much as a train rolls down the tracks. How is a promoter identified? When the holoenzyme encounters a promoter sequence, σ factor recognizes the bases at both the −35 and −10 regions. σ factor protein contains a structure called a **helix-turn-helix motif** that can bind tightly to these regions. Alpha (α) helices within the protein fit into the major groove of the DNA double helix and form hydrogen bonds with the bases. This phenomenon of molecular recognition is shown in **Figure 14.6**. Hydrogen bonding occurs between nucleotides in the −35 and −10 regions of the promoter and amino acid side chains in the helix-turn-helix structure of σ factor.

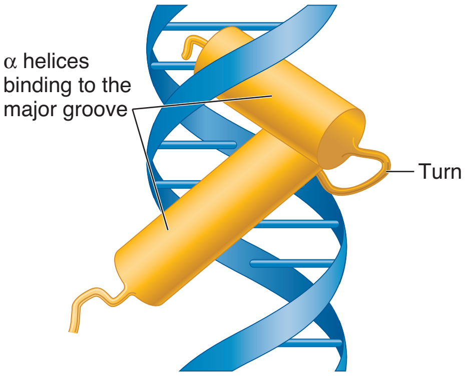

α helices
binding to the
major groove
Turn

FIGURE 14.6 The binding of σ factor protein to the DNA double helix. In this example, the protein contains two α helices connected by a turn, termed a *helix-turn-helix motif.* Two α helices of the protein can fit within the major groove of the DNA. Amino acid side chains within the α helices form hydrogen bonds with the bases in the DNA.

Concept check: Why is it necessary for portions of the sigma (σ) factor protein to fit into the major groove?

As shown in **Figure 14.7**, the process of transcription is initiated when σ factor within the holoenzyme binds to the promoter region to form the **closed complex.** In this closed complex, the strands of DNA in the promoter region are not separated. For transcription to begin, the double-stranded DNA must be unwound to form an open complex. This unwinding first occurs at the TATAAT sequence in the −10 region, which contains only AT base pairs, as shown in Figure 14.4. AT base pairs form only

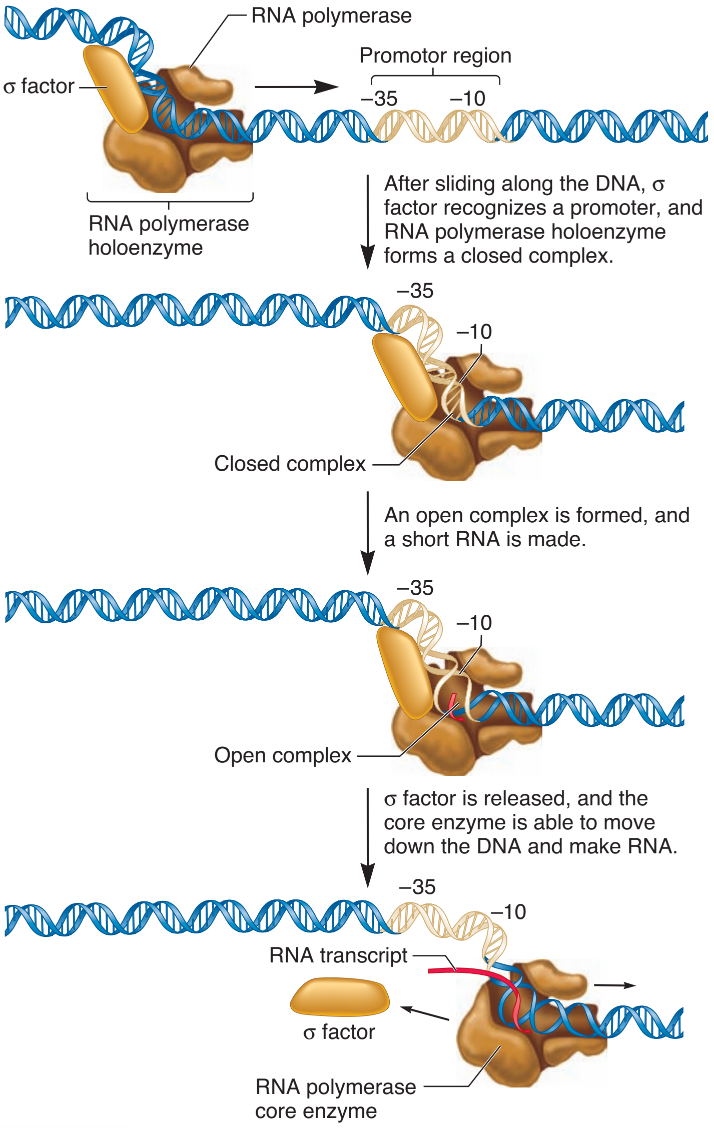

RNA polymerase
Promotor region
σ factor
−35
−10
RNA polymerase
holoenzyme
After sliding along the DNA, σ factor recognizes a promoter, and RNA polymerase holoenzyme forms a closed complex.
−35
−10
Closed complex
An open complex is formed, and a short RNA is made.
−35
−10
Open complex
σ factor is released, and the core enzyme is able to move down the DNA and make RNA.
−35
−10
RNA transcript
σ factor
RNA polymerase
core enzyme

FIGURE 14.7 The initiation stage of transcription in bacteria. The σ factor subunit of the RNA polymerase holoenzyme recognizes the −35 and −10 regions of the promoter. The DNA unwinds in the −10 region to form an open complex, and a short RNA is made. σ factor then dissociates from the holoenzyme, and the RNA polymerase core enzyme can proceed down the DNA to synthesize RNA, forming an open complex as it goes.

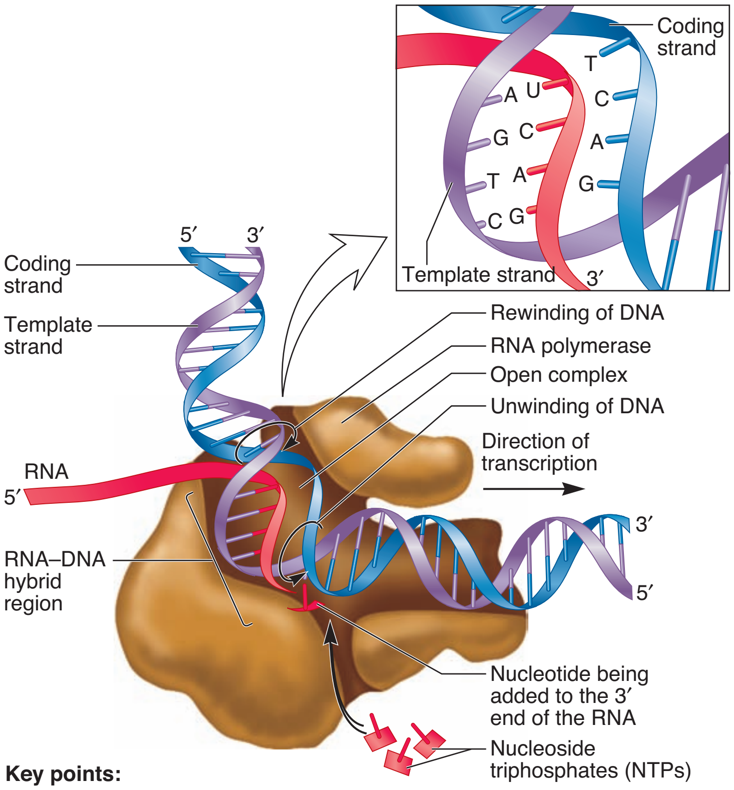
Concept check: What feature of the −10 sequence makes it easy to unwind?

two hydrogen bonds, whereas GC pairs form three. Therefore, DNA in an AT-rich region is more easily separated because fewer hydrogen bonds must be broken. A short strand of RNA is made within the open complex, and then σ factor is released from the core enzyme. The release of σ factor marks the transition to the elongation phase of transcription. The core enzyme may now slide down the DNA to synthesize a strand of RNA.

The RNA Transcript Is Synthesized During the Elongation Stage

After the initiation stage of transcription is completed, the RNA transcript is made during the elongation stage. During the synthesis of the RNA transcript, RNA polymerase moves along the DNA, causing it to unwind (**Figure 14.8**). The DNA strand used as a template for RNA synthesis is termed the **template strand** (also called the antisense strand). The opposite DNA strand is the **coding strand** (also called the sense strand); it has the same sequence as the RNA transcript being synthesized except that T in the DNA corresponds to U in the RNA. Within a given gene, only the template strand is used for RNA synthesis, whereas the coding strand is never used. As it moves along the DNA, the open complex formed by the action of RNA polymerase is approximately 17 bp long. On average, the rate of RNA synthesis is about 43 nucleotides per second! Behind the open complex, the DNA rewinds back into a double helix.

As described in Figure 14.8, the chemistry of transcription by RNA polymerase is similar to the synthesis of DNA via DNA polymerase, which is discussed in Chapter 13. RNA polymerase always connects nucleotides in the 5′ to 3′ direction. During this process, RNA polymerase catalyzes the formation of a bond between the 5′ phosphate group on one nucleotide and the 3′–OH group on the previous nucleotide. The complementarity rule is similar to the AT/GC rule, except that uracil substitutes for thymine in the RNA. In other words, RNA synthesis obeys an $A_{DNA}U_{RNA}$/ $T_{DNA}A_{RNA}$/$G_{DNA}C_{RNA}$/$C_{DNA}G_{RNA}$ rule.


Coding strand
T
C
A
G
A U
G C
T A
C G
Template strand
3′
5′
3′
Coding strand
Template strand
RNA
5′
RNA–DNA hybrid region
Rewinding of DNA
RNA polymerase
Open complex
Unwinding of DNA
Direction of transcription
3′
5′
Nucleotide being added to the 3′ end of the RNA
Nucleoside triphosphates (NTPs)

Key points:

- RNA polymerase slides along the DNA, creating an open complex as it moves.
- The DNA strand known as the template strand is used to make a complementary copy of RNA as an RNA–DNA hybrid.
- RNA polymerase moves along the template strand in a 3′ to 5′ direction, and RNA is synthesized in a 5′ to 3′ direction using nucleoside triphosphates as precursors. Pyrophosphate is released (not shown).
- The complementarity rule is the same as the AT/GC rule except that U is substituted for T in the RNA.

ONLINE ANIMATION

FIGURE 14.8 Synthesis of the RNA transcript.

When considering the transcription of multiple genes within a chromosome, the direction of transcription and the DNA strand used as a template varies among different genes. **Figure 14.9** shows three genes adjacent to each other within a chromosome. Genes *A* and *B* are transcribed from left to right, using the bottom DNA strand as a template. By comparison, gene *C* is transcribed from right to left and uses the top DNA strand as a template. Note that in all three cases, the template strand is read in the 3′ to 5′ direction, and the synthesis of the RNA transcript occurs in a 5′ to 3′ direction.

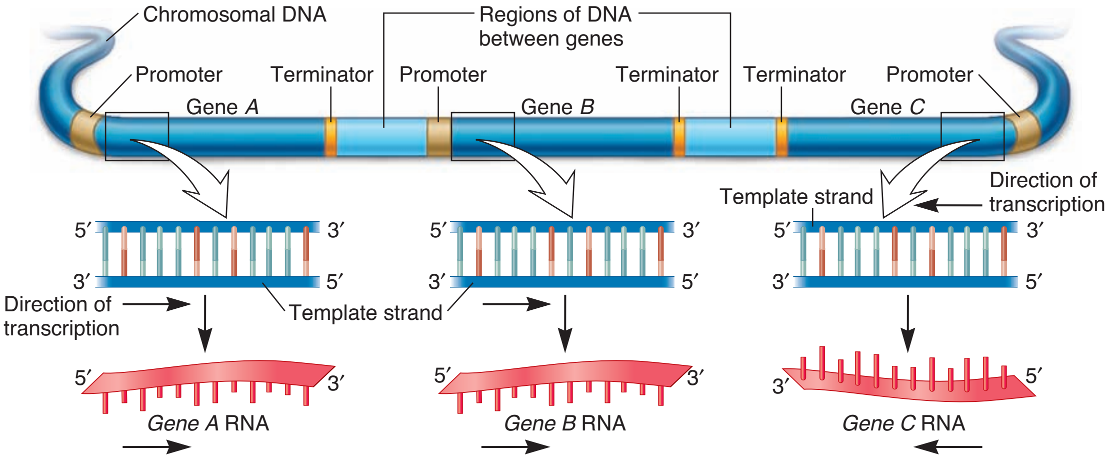
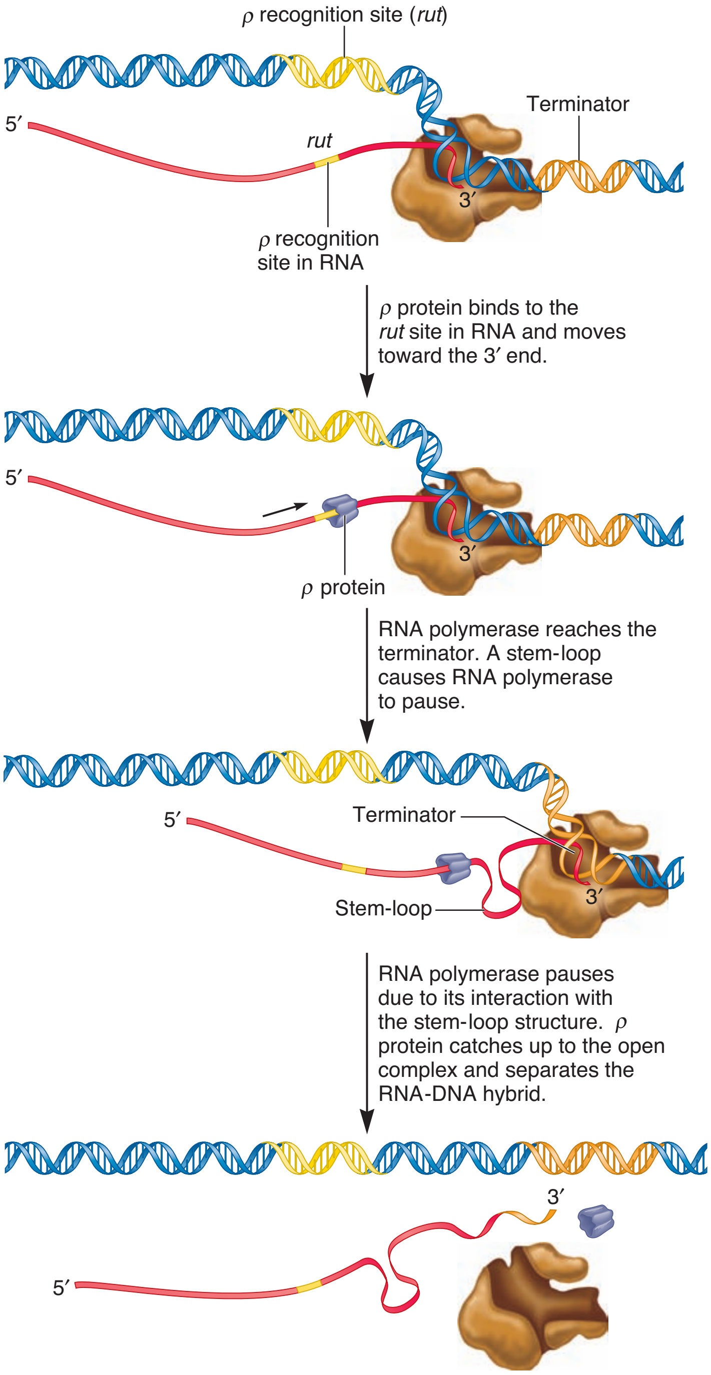

Chromosomal DNA
Regions of DNA between genes
Promoter
Terminator
Promoter
Terminator
Terminator
Promoter
Gene A
Gene B
Gene C
Template strand
Direction of transcription
5′
3′
3′
5′
5′
3′
3′
5′
5′
3′
3′
5′
Direction of transcription
Template strand
5′
3′
Gene A RNA
5′
3′
Gene B RNA
3′
5′
Gene C RNA

FIGURE 14.9 The transcription of three different genes found in the same chromosome. RNA polymerase synthesizes each RNA transcript in a 5′ to 3′ direction, sliding along a DNA template strand in a 3′ to 5′ direction. However, the use of the template strand varies from gene to gene. For example, genes *A* and *B* use the bottom strand, but gene *C* uses the top strand.

Transcription Is Terminated by Either an RNA-Binding Protein or an Intrinsic Terminator

The end of RNA synthesis is referred to as termination. Prior to termination, the hydrogen bonding between the DNA and RNA within the open complex is of central importance in preventing dissociation of RNA polymerase from the template strand. Termination occurs when this short RNA–DNA hybrid region is forced to separate, thereby releasing RNA polymerase as well as the newly made RNA transcript. In *E. coli*, two different mechanisms for termination have been identified. For certain genes, an RNA-binding protein known as **ρ (rho)** is responsible for terminating transcription, in a mechanism called ρ-dependent termination. For other genes, termination does not require the involvement of the ρ protein. This is referred to as ρ-independent termination.

In **ρ-dependent termination,** the termination process requires two components. First, a sequence upstream from the terminator, called the *rut* site for rho utilization site, acts as a recognition site for the binding of the ρ protein (**Figure 14.10**). How does ρ protein facilitate termination? The ρ protein functions as a helicase, an enzyme that can separate RNA–DNA hybrid regions. After the *rut* site is synthesized in the RNA, ρ protein binds to the RNA and moves in the direction of RNA polymerase. The second component of ρ-dependent termination is the site where termination actually takes place. At this terminator site, the DNA encodes an RNA sequence containing several GC base pairs that form a stem-loop. RNA synthesis terminates several nucleotides beyond this stem-loop. As discussed in Chapter 11, a stem-loop, also called a hairpin, can form due to complementary sequences within the RNA (refer back to Figure 11.18). This stem-loop forms almost immediately after the RNA sequence is synthesized and quickly binds to RNA polymerase. This binding results in a conformational change that causes RNA polymerase to pause in its synthesis of RNA. The pause allows ρ protein to catch up to the stem-loop, pass through it, and break the hydrogen bonds between the DNA and RNA within the open complex. When this occurs, the completed RNA strand is separated from the DNA along with RNA polymerase.


ρ recognition site (*rut*)
5′
rut
Terminator
3′
ρ recognition site in RNA
ρ protein binds to the *rut* site in RNA and moves toward the 3′ end.
5′
3′
ρ protein
RNA polymerase reaches the terminator. A stem-loop causes RNA polymerase to pause.
5′
Terminator
Stem-loop
3′
RNA polymerase pauses due to its interaction with the stem-loop structure. ρ protein catches up to the open complex and separates the RNA-DNA hybrid.
3′
5′

ONLINE ANIMATION

FIGURE 14.10 **ρ-Dependent termination.**

Concept check: What would be the consequences if a mutation removed the rut site from this RNA molecule?

Let's now turn our attention to **ρ-independent termination,** a process that does not require the ρ protein. In this case, the terminator is composed of two adjacent nucleotide sequences that function within the RNA (**Figure 14.11**). One is a uracil-rich sequence located at the 3′ end of the RNA. The second sequence is adjacent to the uracil-rich sequence and promotes the formation of a stem-loop. As shown in Figure 14.11, the formation of the stem-loop causes RNA polymerase to pause in its synthesis of RNA. This pausing is stabilized by other proteins that bind to RNA polymerase. For example, a protein called NusA, which is bound to RNA polymerase, promotes pausing at stem-loops. At the precise time RNA polymerase pauses, the uracil-rich sequence in the RNA transcript is bound to the DNA template strand. As previously mentioned, the hydrogen bonding of RNA to DNA keeps RNA polymerase clamped onto the DNA. However, the binding of this uracil-rich sequence to the DNA template strand is relatively weak, causing the RNA transcript to spontaneously dissociate from the DNA and cease further transcription. Because this process does not require a protein (the ρ protein) to physically remove the RNA transcript from the DNA, it is also referred to as **intrinsic termination.** In *E. coli*, about half of the genes show intrinsic termination, and the other half are terminated by ρ protein.

14.2 REVIEWING THE KEY CONCEPTS

- Many bacterial promoters have sequence elements at the −35 and −10 regions. The transcriptional start site is at +1 (see Figures 14.4, 14.5).

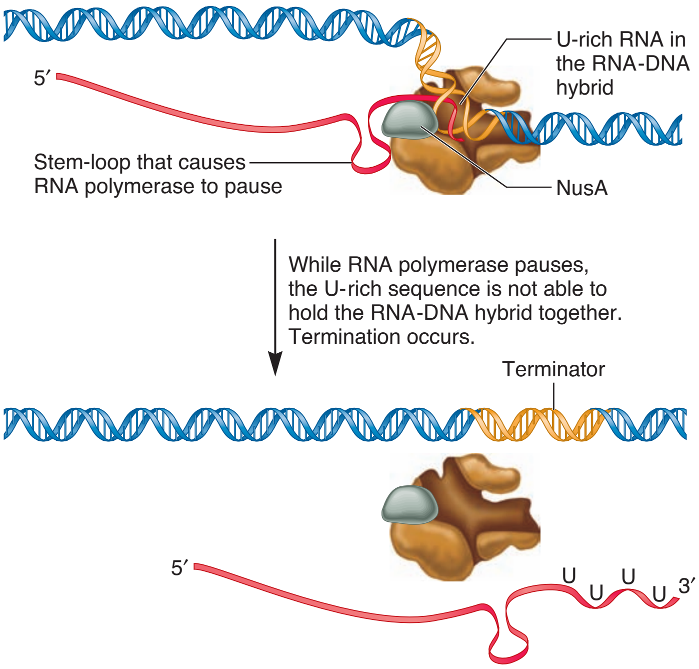

U-rich RNA in the RNA-DNA hybrid
5′
Stem-loop that causes RNA polymerase to pause
NusA
While RNA polymerase pauses, the U-rich sequence is not able to hold the RNA-DNA hybrid together. Termination occurs.
Terminator
5′
U
U
U
U
3′

FIGURE 14.11 **ρ-Independent or intrinsic termination.** When RNA polymerase reaches the end of the gene, it transcribes a uracil-rich sequence. As this uracil-rich sequence is transcribed, a stem-loop forms just upstream from the open complex. The formation of this stem-loop causes RNA polymerase to pause in its synthesis of the transcript. This pausing is stabilized by NusA, which binds near the region where RNA exits the open complex. While it is pausing, the RNA in the RNA-DNA hybrid is a uracil-rich sequence. Because hydrogen bonds between U and A are relatively weak interactions, the transcript and RNA polymerase dissociate from the DNA.

Concept check: Why is NusA important for this process?

- During the initiation phase of transcription in *E. coli*, sigma (σ) factor, which is bound to RNA polymerase, binds into the major groove of DNA and recognizes sequence elements at the promoter. This process forms a closed complex. Following the formation of an open complex, σ factor is released (see Figures 14.6, 14.7).
- During the elongation phase of transcription, RNA polymerase slides along the DNA and maintains an open complex as it goes. RNA is made in the 5′ to 3′ direction according to complementary base pairing (see Figure 14.8).
- In a given chromosome, the use of the template strand varies from gene to gene (see Figure 14.9).
- Transcriptional termination in *E. coli* occurs by a rho (ρ)-dependent or ρ-independent mechanism (see Figures 14.10, 14.11).

14.2 COMPREHENSION QUESTIONS

1. With regard to a promoter, a transcriptional start site is
   a. located at the –35 position and recognized by σ factor.
   b. located at the –35 position and is where the first base is used as a template for RNA transcription.
   c. located at the +1 position and recognized by σ factor.
   d. located at the +1 position and is where the first base is used as a template for RNA transcription.
2. For the following five sequences, what is the consensus sequence?
   5′-GGGAGCG-3′
   5′-GAGAGCG-3′
   5′-GAGTGCG-3′
   5′-GAGAACG-3′
   5′-GAGAGCA-3′
   a. 5′-GGGAGCG-3′
   b. 5′-GAGAGCG-3′
   c. 5′-GAGTGCG-3′
   d. 5′-GAGAACG-3′
3. Sigma factor (σ) is needed during which stage(s) of transcription?
   a. Initiation
   b. Elongation
   c. Termination
   d. All of the above
4. A uracil-rich sequence occurs at the end of the RNA in
   a. ρ-dependent termination.
   b. ρ-independent termination.
   c. both a and b.
   d. none of the above.

14.3 TRANSCRIPTION IN EUKARYOTES

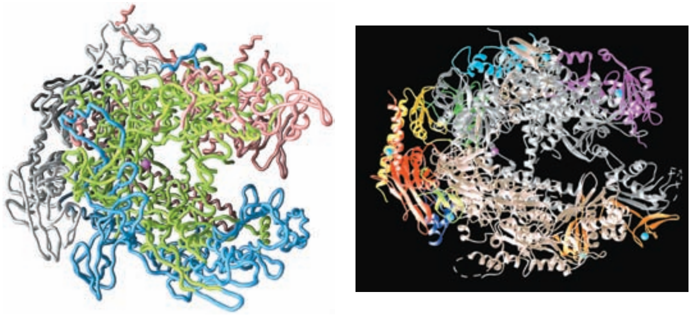
Many of the basic features of gene transcription are very similar in bacterial and eukaryotic species. Much of our understanding of transcription has come from studies in *Saccharomyces cerevisiae* (baker's yeast) and other eukaryotic species, including mammals. In general, gene transcription in eukaryotes is more complex than that of their bacterial counterparts. Eukaryotic cells are larger and contain a variety of compartments known as organelles. This added level of cellular complexity dictates that eukaryotes contain many more genes encoding cellular proteins. In addition, most eukaryotic species are multicellular, being composed of many different cell types. Multicellularity adds the requirement that genes be transcribed in the correct type of cell and during the proper stage of development. Therefore, in any given species, the transcription of the thousands of different genes that an organism possesses requires appropriate timing and coordination. In this section, we will examine the basic features of gene transcription in eukaryotes. We will focus on the proteins that are needed to make an RNA transcript. In addition, an important factor that affects eukaryotic gene transcription is chromatin structure. Eukaryotic gene transcription requires changes in the positions and structures of nucleosomes. However, because these changes are important for regulating transcription, they are described in Chapter 17 rather than in this chapter.

Eukaryotes Have Multiple RNA Polymerases That Are Structurally Similar to the Bacterial Enzyme

The genetic material within the nucleus of a eukaryotic cell is transcribed by three different RNA polymerase enzymes—designated RNA polymerase I, II, and III. What are the roles of these enzymes?

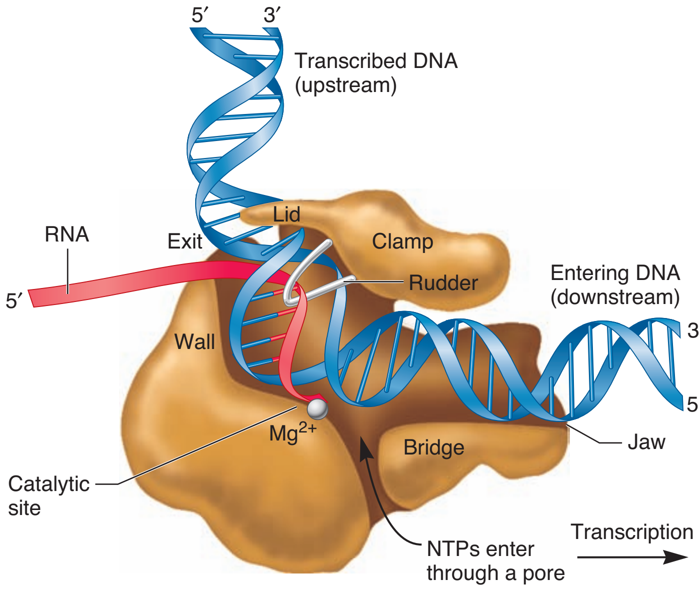
Each of the three RNA polymerases transcribes different categories of genes. RNA polymerase I transcribes all of the genes that encode ribosomal RNA (rRNA), except for the 5S rRNA. RNA polymerase II plays a major role in transcription because it transcribes all of the structural genes. It is responsible for the synthesis of all mRNA and also transcribes certain snRNA genes, which are needed for pre-mRNA splicing that is described later in this chapter. RNA polymerase III transcribes all tRNA genes and the 5S rRNA gene.

All three RNA polymerases are structurally very similar to each other and are composed of many subunits. They contain two large catalytic subunits similar to the β and β′ subunits of bacterial RNA polymerase. The structures of RNA polymerase from a few different species have been determined by X-ray crystallography. A remarkable similarity also exists between the bacterial enzyme and its eukaryotic counterparts. Figure 14.12a compares the structures of a bacterial RNA polymerase with RNA polymerase II from yeast. As seen here, both enzymes have a very similar structure. Also, it is very exciting that this structure provides a way to envision how the transcription process works. As seen in Figure 14.12b, DNA enters the enzyme through the jaw and lies on a surface within RNA polymerase termed the bridge. The part of the enzyme called the clamp is thought to control the movement of the DNA through RNA polymerase. A wall in the enzyme forces the RNA-DNA hybrid to make a right-angle turn. This bend facilitates the ability of nucleotides to bind to the template strand. Mg^{2+} is located at the catalytic site, which is precisely at the 3′ end of the growing RNA strand. Nucleoside triphosphates (NTPs) enter the catalytic site via a pore region. The correct nucleotide binds to the template DNA and is covalently attached to the 3′ end. As RNA polymerase slides down the template, a rudder, which is about 9 bp away from the 3′ end of the RNA, forces the RNA-DNA hybrid apart. The single-stranded RNA then exits under a small lid.



(a) Structure of a bacterial RNA polymerase

Structure of a eukaryotic RNA polymerase II (yeast)


5′
3′
Transcribed DNA (upstream)
RNA
Exit
Lid
Clamp
5′
Rudder
Entering DNA (downstream)
3′
Wall
5′
Mg2+
Jaw
Bridge
Catalytic site
Transcription
NTPs enter through a pore

(b) Schematic structure of RNA polymerase

FIGURE 14.12 **Structure and molecular function of RNA polymerase.** **(a)** A comparison of the crystal structures of a bacterial RNA polymerase (left) to a eukaryotic RNA polymerase II (right). The bacterial enzyme is from *Thermus aquaticus*. The eukaryotic enzyme is from *Saccharomyces cerevisiae.* **(b)** A mechanism for transcription based on the crystal structure. In this diagram, the direction of transcription is from left to right. The double-stranded DNA enters the polymerase along a bridge surface that is between the jaw and clamp. At a region termed the wall, the RNA-DNA hybrid is forced to make a right-angle turn, which enables nucleotides to bind to the template strand. Mg^{2+} is located at the catalytic site. Nucleoside triphosphates (NTPs) enter the catalytic site via a pore region and bind to the template DNA. At the catalytic site, the nucleotides are covalently attached to the 3′ end of the RNA. As RNA polymerase slides down the template, a small region of the protein termed the rudder separates the RNA-DNA hybrid. The single-stranded RNA then exits under a small lid.

Eukaryotic Structural Genes Have a Core Promoter and Regulatory Elements

In eukaryotes, the promoter sequence is more variable and often more complex than that found in bacteria. For structural genes, at least three features are found in most promoters: regulatory elements, a TATA box, and a transcriptional start site. Figure 14.13 shows a common pattern of sequences found within the promoters of eukaryotic structural genes. The **core promoter** is a relatively short DNA sequence that is necessary for transcription to take place. It consists of a TATAAA sequence called the **TATA box** and the transcriptional start site, where transcription begins. The TATA box, which is usually about 25 bp upstream from a transcriptional start site, is important in determining the precise starting point for transcription. If it is missing from the core promoter, the transcription start site becomes undefined, and transcription may start at a variety of different locations. The core promoter, by itself, produces a low level of transcription. This is termed **basal transcription.**

Regulatory elements are short DNA sequences that affect the ability of RNA polymerase to recognize the core promoter and begin the process of transcription. These elements are recognized by transcription factors—proteins that bind to regulatory elements and influence the rate of transcription. There are two categories of regulatory elements. Activating sequences, known as **enhancers,** are needed to stimulate transcription. In the absence of enhancer sequences, most eukaryotic genes have very low levels of basal transcription. Under certain conditions, it may also be necessary to prevent transcription of a given gene. This occurs

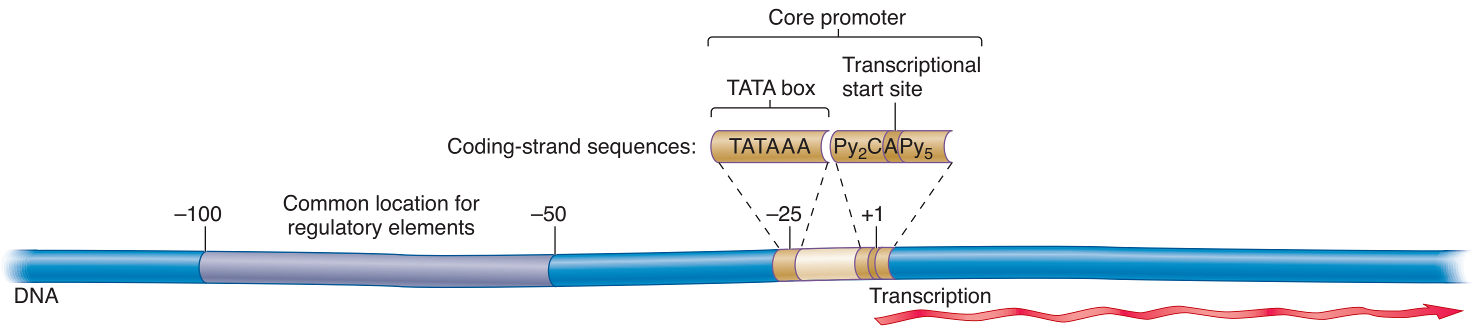


FIGURE 14.13 **A common pattern found for the promoter of structural genes recognized by RNA polymerase II.** The start site usually occurs at adenine; two pyrimidines (Py; cytosine or thymine) and a cytosine precede this adenine, and five pyrimidines (Py) follow it. A TATA box is approximately 25 bp upstream. However, the sequences that constitute eukaryotic promoters are quite diverse, and not all structural genes have a TATA box. Regulatory elements, which are discussed in Chapter 17, vary in their locations but are often found in the –50 to –100 region. The core promoters for RNA polymerase I and III are quite different. A single upstream regulatory element is involved in the binding of RNA polymerase I to its promoter, whereas two regulatory elements, called A and B boxes, facilitate the binding of RNA polymerase III.

Concept check: *What is the functional role of the TATA box?*

via **silencers**—DNA sequences that inhibit transcription. As seen in Figure 14.13, a common location for regulatory elements is the −50 to −100 region. However, the locations of regulatory elements vary considerably among different eukaryotic genes. These elements can be far away from the core promoter yet strongly influence the ability of RNA polymerase to initiate transcription.

DNA sequences such as the TATA box, enhancers, and silencers exert their effects only over a particular gene. They are called ***cis*-acting elements.** The term *cis* comes from chemistry nomenclature meaning "next to." *Cis*-acting elements, though possibly far away from the core promoter, are always found within the same chromosome as the genes they regulate. By comparison, the transcription factors that bind to such elements are called ***trans*-acting factors** (the term *trans* means "across from"). The transcription factors that control the expression of a gene are themselves encoded by genes; regulatory genes that encode transcription factors may be far away from the genes they control. When a gene encoding a *trans*-acting factor is expressed, the transcription factor protein that is made can diffuse throughout the cell and bind to its appropriate *cis*-acting element. Let's now turn our attention to the function of such proteins.

Transcription of Eukaryotic Structural Genes Is Initiated When RNA Polymerase II and General Transcription Factors Bind to a Promoter Sequence

Thus far, we have considered the DNA sequences that play a role in the promoter region of eukaryotic structural genes. By studying transcription in a variety of eukaryotic species, researchers have discovered that three categories of proteins are needed for basal transcription at the core promoter: RNA polymerase II, general transcription factors, and a complex called mediator (Table 14.2).

Five different proteins called **general transcription factors (GTFs)** are always needed for RNA polymerase II to initiate transcription of structural genes. Figure 14.14 describes the assembly of GTFs and RNA polymerase II at the TATA box. As

TABLE **14.2**

Proteins Needed for Transcription via the Core Promoter of Eukaryotic Structural Genes

RNA polymerase II: The enzyme that catalyzes the linkage of ribonucleotides in the 5′ to 3′ direction, using DNA as a template. Essentially all eukaryotic RNA polymerase II proteins are composed of 12 subunits. The two largest subunits are structurally similar to the β and β′ subunits found in *E. coli* RNA polymerase.

General transcription factors:

TFIID:	Composed of TATA-binding protein (TBP) and other TBP-associated factors (TAFs). Recognizes the TATA box of eukaryotic structural gene promoters.
TFIIB:	Binds to TFIID and then enables RNA polymerase II to bind to the core promoter. Also promotes TFIIF binding.
TFIIF:	Binds to RNA polymerase II and plays a role in its ability to bind to TFIIB and the core promoter. Also plays a role in the ability of TFIIE and TFIIH to bind to RNA polymerase II.
TFIIE:	Plays a role in the formation or the maintenance (or both) of the open complex. It may exert its effects by facilitating the binding of TFIIH to RNA polymerase II and regulating the activity of TFIIH.
TFIIH:	A multisubunit protein that has multiple roles. First, certain subunits act as helicases and promote the formation of the open complex. Other subunits phosphorylate the carboxyl terminal domain (CTD) of RNA polymerase II, which releases its interaction with TFIIB, thereby allowing RNA polymerase II to proceed to the elongation phase.

Mediator: A multisubunit complex that mediates the effects of regulatory transcription factors on the function of RNA polymerase II. Though mediator typically has certain core subunits, many of its subunits vary, depending on the cell type and environmental conditions. The ability of mediator to affect the function of RNA polymerase II is thought to occur via the CTD of RNA polymerase II. Mediator can influence the ability of TFIIH to phosphorylate CTD, and subunits within mediator itself have the ability to phosphorylate CTD. Because CTD phosphorylation is needed to release RNA polymerase II from TFIIB, mediator plays a key role in the ability of RNA polymerase II to switch from the initiation to the elongation stage of transcription.

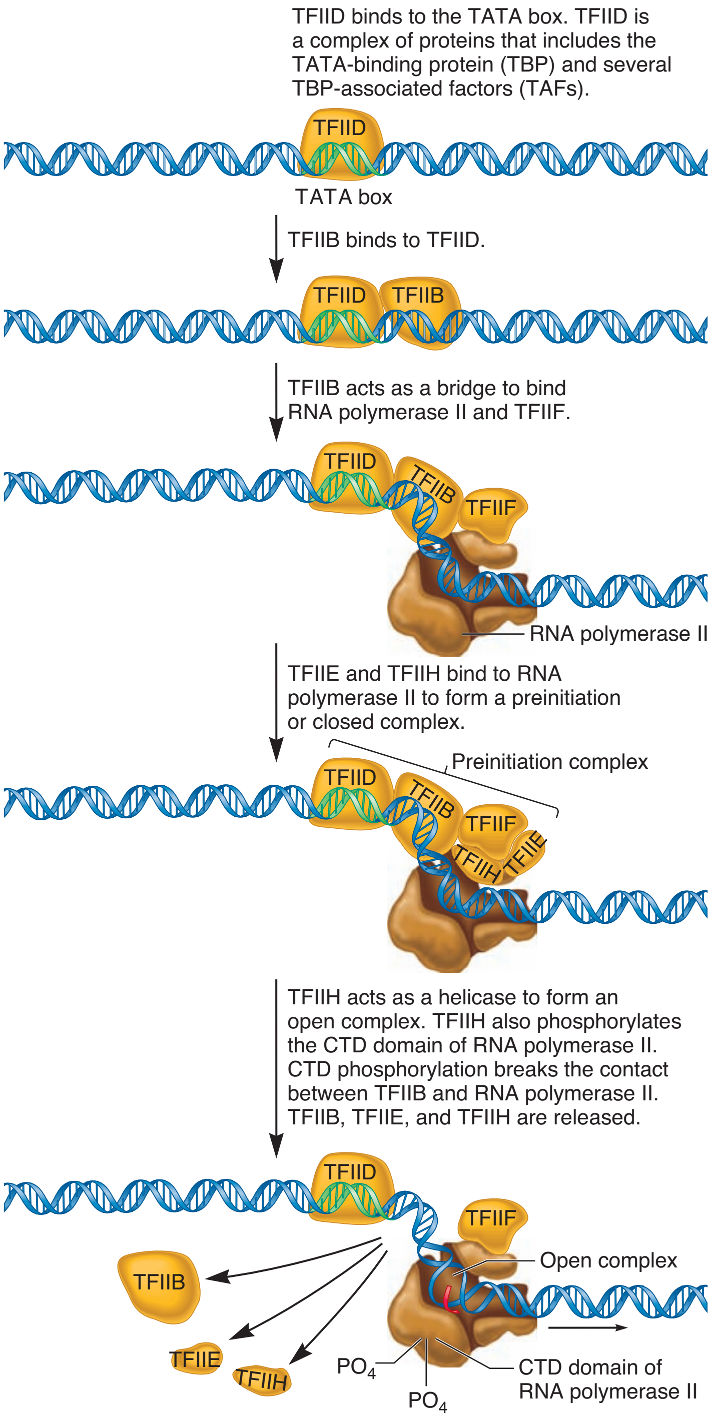

TFIID binds to the TATA box. TFIID is a complex of proteins that includes the TATA-binding protein (TBP) and several TBP-associated factors (TAFs).
TFIID
TATA box
TFIIB binds to TFIID.
TFIID
TFIIB
TFIIB acts as a bridge to bind RNA polymerase II and TFIIF.
TFIID
TFIIB
TFIIF
RNA polymerase II
TFIIE and TFIIH bind to RNA polymerase II to form a preinitiation or closed complex.
Preinitiation complex
TFIID
TFIIB
TFIIF
TFIIH
TFIIE
TFIIH acts as a helicase to form an open complex. TFIIH also phosphorylates the CTD domain of RNA polymerase II. CTD phosphorylation breaks the contact between TFIIB and RNA polymerase II. TFIIB, TFIIE, and TFIIH are released.
TFIID
TFIIF
Open complex
TFIIB
TFIIE
TFIIH
PO4
PO4
CTD domain of RNA polymerase II

FIGURE 14.14 Steps leading to the formation of the open complex. The DNA is first denatured at the TATA box to form an open complex. In this diagram, the open complex has moved to the transcriptional start site, which is usually about 25 bp away from the TATA box. Note: Though evidence suggests that TFIIH is often released during the early stage of elongation, there may be circumstances in which it is not released and plays a role in elongation.

Concept check: Why is carboxyl terminal domain (CTD) phosphorylation functionally important?

shown here, a series of interactions leads to the formation of the open complex. Transcription factor IID (TFIID) first binds to the TATA box, thereby playing a critical role in the recognition of the core promoter. TFIID is composed of several subunits, including TATA-binding protein (TBP), which directly binds to the TATA box, and several other proteins called TBP-associated factors (TAFs). After TFIID binds to the TATA box, it associates with TFIIB. TFIIB promotes the binding of RNA polymerase II and TFIIF to the core promoter. Lastly, TFIIE and TFIIH bind to the complex. This completes the assembly of proteins to form a closed complex, also known as a **preinitiation complex.**

TFIIH plays a major role in the formation of the open complex. TFIIH has several subunits that perform different functions. Certain subunits act as helicases, which break the hydrogen bonding between the double-stranded DNA and thereby promote the formation of the open complex. Another subunit hydrolyzes ATP and phosphorylates a domain in RNA polymerase II known as the carboxyl terminal domain (CTD). Phosphorylation of the CTD releases the contact between RNA polymerase II and TFIIB. Next, TFIIB, TFIIE, and TFIIH dissociate, and RNA polymerase II is free to proceed to the elongation stage of transcription.

In vitro, when researchers mix together TFIID, TFIIB, TFIIF, TFIIE, TFIIH, RNA polymerase II, and a DNA sequence containing a TATA box and transcriptional start site, the DNA is transcribed into RNA. Therefore, these components are referred to as the **basal transcription apparatus.** In a living cell, however, additional components regulate transcription and allow it to proceed at a reasonable rate.

In addition to GTFs and RNA polymerase II, another component required for transcription is a large protein complex termed mediator. This complex was discovered by Roger Kornberg and colleagues in 1990. In 2006, Kornberg was awarded the Nobel Prize in chemistry for his studies regarding the molecular basis of eukaryotic transcription. **Mediator** derives its name from the observation that it mediates interactions between RNA polymerase II and regulatory transcription factors that bind to enhancers or silencers. It serves as an interface between RNA polymerase II and many, diverse regulatory signals. The subunit composition of mediator is quite complex and variable. The core subunits form an elliptically shaped complex that partially wraps around RNA polymerase II. Mediator itself may phosphorylate the CTD of RNA polymerase II, and it may regulate the ability of TFIIH to phosphorylate the CTD. Therefore, it plays a pivotal role in the switch between transcriptional initiation and elongation.

Transcriptional Termination of RNA Polymerase II Occurs After the 3′ End of the Transcript Is Cleaved Near the PolyA Signal Sequence

As discussed later in this chapter, eukaryotic mRNAs are modified by cleavage near their 3′ end and the subsequent attachment of a string of adenine nucleotides (look ahead at Figure 14.22). This process, which is called polyadenylation, requires a polyA signal sequence that directs the cleavage of the pre-mRNA. Transcription via RNA polymerase II typically terminates about 500 to 2000 nucleotides downstream from the polyA signal.

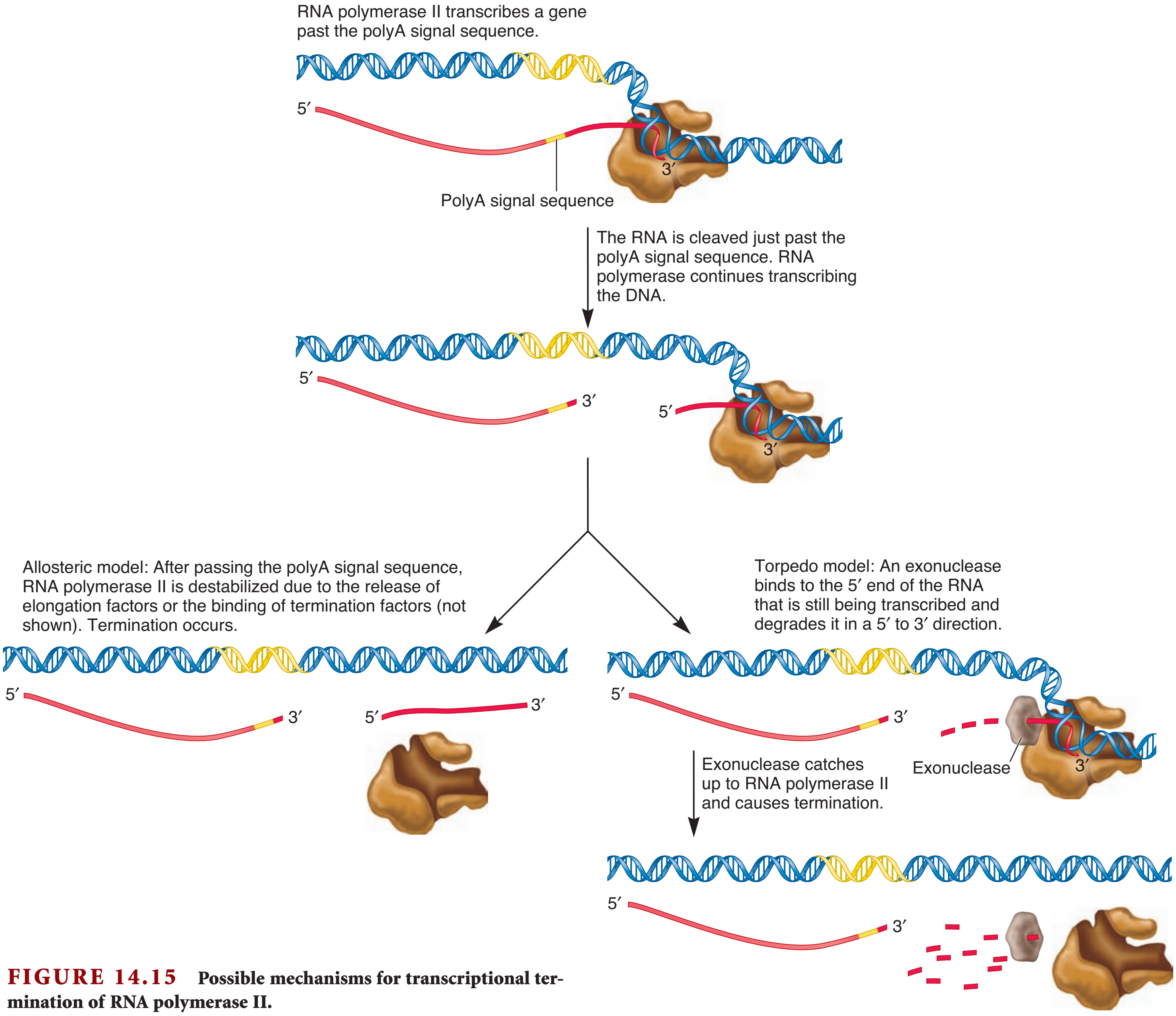

RNA polymerase II transcribes a gene past the polyA signal sequence.
5′
3′
PolyA signal sequence
The RNA is cleaved just past the polyA signal sequence. RNA polymerase continues transcribing the DNA.
5′
3′
5′
3′
Allosteric model: After passing the polyA signal sequence, RNA polymerase II is destabilized due to the release of elongation factors or the binding of termination factors (not shown). Termination occurs.
5′
3′
5′
3′
Torpedo model: An exonuclease binds to the 5′ end of the RNA that is still being transcribed and degrades it in a 5′ to 3′ direction.
5′
3′
Exonuclease
3′
Exonuclease catches up to RNA polymerase II and causes termination.
5′
3′

FIGURE 14.15 **Possible mechanisms for transcriptional termination of RNA polymerase II.**

Figure 14.15 shows a simplified scheme for the transcriptional termination of RNA polymerase II. After RNA polymerase II has transcribed the polyA signal sequence, the RNA is cleaved just downstream from this sequence. This cleavage occurs before transcriptional termination. Two models have been proposed for transcriptional termination. According to the allosteric model, RNA polymerase II becomes destabilized after it has transcribed the polyA signal sequence, and it eventually dissociates from the DNA. This destabilization may be caused by the loss of proteins that function as elongation factors or by the binding of proteins that function as termination factors. A second model, called the torpedo model, suggests that RNA polymerase II is physically removed from the DNA. According to this model, the region of RNA that is downstream from the polyA signal sequence is cleaved by an exonuclease that degrades the transcript in the 5′ to 3′ direction. When the exonuclease catches up to RNA polymerase II, this causes RNA polymerase II to dissociate from the DNA.

Which of these two models is correct? Additional research is needed, but the results of studies over the past few years have provided evidence that the two models are not mutually exclusive. Therefore, both mechanisms may play a role in transcriptional termination.

14.3 REVIEWING THE KEY CONCEPTS

- Eukaryotes use RNA polymerase I, II, and III to transcribe different categories of genes. Prokaryotic and eukaryotic RNA polymerases have similar structures (see Figure 14.12).
- Eukaryotic promoters have a core promoter and regulatory elements such as enhancers and silencers (see Figure 14.13).
- Transcription of structural genes in eukaryotes requires RNA polymerase II, five general transcription factors, and mediator.

The five general transcription factors and RNA polymerase assemble together to form an open complex (see Table 14.2, Figure 14.14).

- Transcriptional termination of RNA polymerase II may occur via the allosteric or torpedo model (see Figure 14.15).

14.3 COMPREHENSION QUESTIONS

1. Which RNA polymerase in eukaryotes is responsible for the transcription of genes that encode proteins?
   a. RNA polymerase I
   b. RNA polymerase II
   c. RNA polymerase III
   d. All of the above
2. An enhancer is a ______________ that ____________ the rate of transcription.
   a. *trans*-acting factor, increases
   b. *trans*-acting factor, decreases
   c. *cis*-acting element, increases
   d. *cis*-acting element, decreases
3. The basal transcription apparatus is composed of
   a. five general transcription factors.
   b. RNA polymerase II.
   c. a DNA sequence containing a TATA box and transcriptional start site.
   d. all of the above.
4. With regard to transcriptional termination in eukaryotes, which model suggests that RNA polymerase is physically removed from the DNA?
   a. Allosteric model
   b. Torpedo model
   c. Both models
   d. Neither model

14.4 RNA MODIFICATION

During the 1960s and 1970s, studies in bacteria established the physical structure of the gene. The analysis of bacterial genes showed that the sequence of DNA within the coding strand corresponds to the sequence of nucleotides in the mRNA, except that T is replaced with U. During translation, the sequence of codons in the mRNA is then read, providing the instructions for the correct amino acid sequence in a polypeptide. The one-to-one correspondence between the sequence of codons in the DNA-coding strand and the amino acid sequence of the polypeptide has been termed the **colinearity** of gene expression.

The situation dramatically changed in the late 1970s, when the tools became available to study eukaryotic genes at the molecular level. The scientific community was astonished by the discovery that eukaryotic structural genes are not always colinear with their functional mRNAs. Instead, the coding sequences within many eukaryotic genes are separated by DNA sequences that are not translated into protein. The coding sequences are found within **exons,** which are regions that are contained within mature RNA. By comparison, the sequences that are found between the exons are called **intervening sequences,** or **introns.** During transcription, an RNA is made corresponding to the entire gene sequence. Subsequently, as it matures, the sequences in the RNA that correspond to the introns are removed and the exons are connected, or spliced, together. This process is called **RNA splicing.** Since the 1970s, research has revealed that splicing is a common genetic phenomenon in eukaryotic species. Splicing occurs occasionally in bacteria as well.

Aside from splicing, research has also shown that RNA transcripts can be modified in several other ways. **Table 14.3** describes the general types of RNA modifications. For example, rRNAs and tRNAs are synthesized as long transcripts that are processed into smaller functional pieces. In addition, most eukaryotic mRNAs have a cap attached to their 5′ end and a tail attached at their 3′ end. In this section, we will examine the molecular mechanisms that account for several types of RNA modifications and consider why they are functionally important.

Some Large RNA Transcripts Are Cleaved into Smaller Functional Transcripts

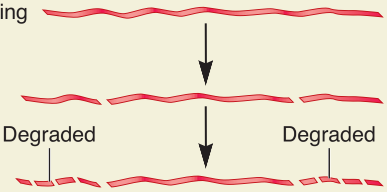
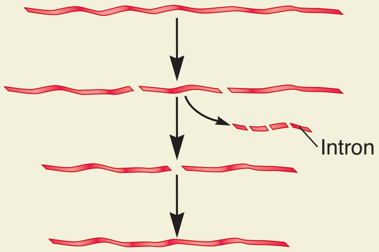
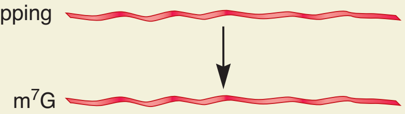
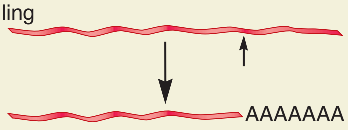
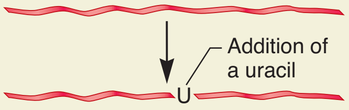
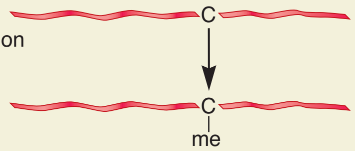
For many nonstructural genes, the RNA transcript initially made during gene transcription is processed or cleaved into smaller pieces. As an example, **Figure 14.16** shows the processing of mammalian ribosomal RNA (rRNA). The rRNA gene is transcribed by RNA polymerase I to make a long primary transcript, known as 45S rRNA. The term 45S refers to the sedimentation characteristics of this transcript in Svedberg units. Following the synthesis of the 45S rRNA, cleavage occurs at several points to produce three fragments: termed 18S, 5.8S, and 28S rRNA. These are functional rRNA molecules that play a key role in forming the structure of the ribosome. In eukaryotes, the cleavage of 45S rRNA into smaller rRNAs and the assembly of ribosomal subunits occur in a structure within the cell nucleus known as the **nucleolus.**

The production of tRNA molecules requires processing via exonucleases and endonucleases. An **exonuclease** cleaves a covalent bond between two nucleotides at one end of a strand. Starting at one end, an exonuclease digests a strand, one nucleotide at a time. Some exonucleases begin this digestion only from the 3′ end, traveling in the 3′ to 5′ direction, whereas others begin only at the 5′ end and digest in the 5′ to 3′ direction. By comparison, an **endonuclease** cleaves the bond between two adjacent nucleotides within a strand.

Like ribosomal RNA, tRNAs are synthesized as large precursor tRNAs that must be cleaved to produce mature, functional tRNAs that carry amino acids. This processing has been studied extensively in *E. coli*. **Figure 14.17** shows the processing of a precursor tRNA, which involves the action of two endonucleases and one exonuclease. The precursor tRNA is recognized by RNaseP, which is an endonuclease that cuts the precursor tRNA. The action of RNaseP produces the correct 5′ end of the mature tRNA. A different endonuclease cleaves the precursor tRNA to remove a 170-nucleotide segment from the 3′ end. Next, an exonuclease called RNaseD binds to the 3′ end and digests the RNA in the 3′

TABLE 14.3

Modifications That May Occur to RNAs

Modification	Description
Processing 	The cleavage of a large RNA transcript into smaller pieces. One or more of the smaller pieces becomes a functional RNA molecule. Processing occurs for rRNA and tRNA transcripts. Occurrence: Occurs in both prokaryotes and eukaryotes.
Splicing 	Splicing involves both cleavage and joining of RNA molecules. The RNA is cleaved at two sites, which allows an internal segment of RNA, known as an intron, to be removed. After the intron is removed, the two ends of the RNA molecules are joined together. Occurrence: Splicing is common among eukaryotic pre-mRNAs, and it also occurs occasionally in rRNAs, tRNAs, and a few bacterial RNAs.
5′ capping 	The attachment of a 7-methylguanosine cap (m^7G) to the 5′ end of mRNA. The cap plays a role in the splicing of introns, the exit of mRNA from the nucleus, and the binding of mRNA to the ribosome. Occurrence: Capping occurs on eukaryotic mRNAs.
3′ polyA tailing 	The attachment of a string of adenine-containing nucleotides to the 3′ end of mRNA at a site where the mRNA is cleaved (see upward arrow). It is important for RNA stability and translation in eukaryotes. Occurrence: Occurs on eukaryotic mRNAs and occasionally occurs on bacterial RNAs.
RNA editing 	The change of the base sequence of an RNA after it has been transcribed. Occurrence: Occurs occasionally among eukaryotic RNAs.
Base modification 	The covalent modification of a base within an RNA molecule. Occurrence: Base modification commonly occurs in tRNA molecules found in both prokaryotes and eukaryotes. C—me indicates methylation of cytosine.

to 5′ direction. When it reaches an ACC sequence, the exonuclease stops digesting the precursor tRNA molecule. Therefore, all tRNAs in *E. coli* have an ACC sequence at their 3′ ends. Finally, certain bases in tRNA molecules may be covalently modified to alter their structure. The functional importance of modified bases in tRNAs is discussed in Chapter 15.

As researchers studied tRNA processing, they discovered certain catalytic features that were very unusual and exciting, changing the way biologists view the actions of catalysts. RNaseP has been found to contain both RNA and protein subunits. In 1983, Sidney Altman and colleagues made the surprising discovery that the RNA portion of RNaseP, not the protein subunit, contains the catalytic ability to cleave the precursor tRNA. RNaseP is an example of a **ribozyme,** an RNA molecule with catalytic activity. Prior to the study of RNaseP and the identification of self-splicing RNAs (discussed later), biochemists had staunchly believed that only enzymes, which are protein molecules, could function as biological catalysts.

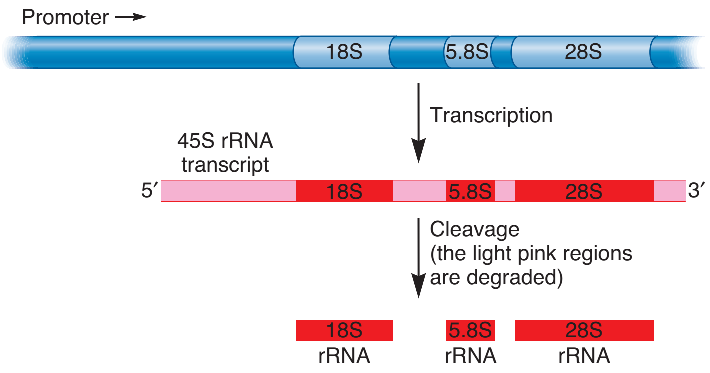


FIGURE 14.16 The processing of ribosomal RNA in eukaryotes. The large ribosomal RNA gene is transcribed into a long 45S rRNA primary transcript. This transcript is cleaved to produce 18S, 5.8S, and 28S rRNA molecules, which become associated with protein subunits in the ribosome. This processing occurs within the nucleolus.

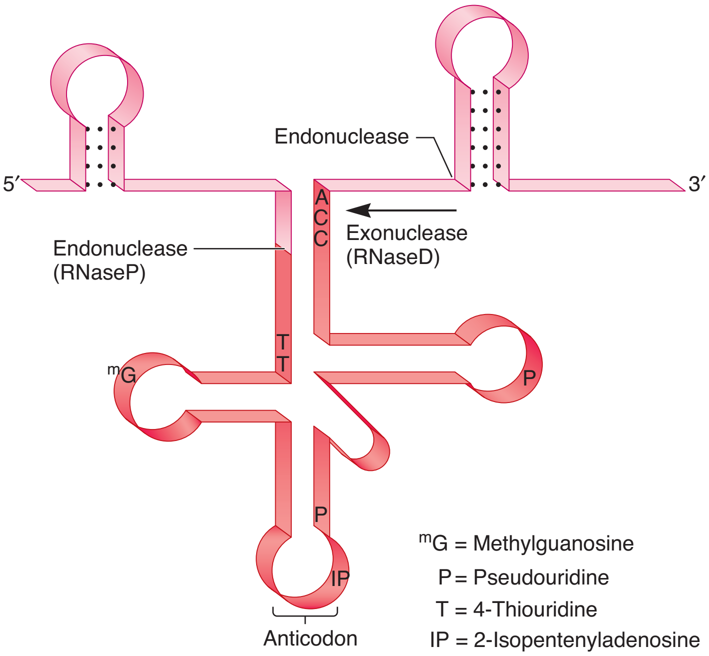


FIGURE 14.17 The processing of a precursor tRNA molecule in *E. coli*. RNaseP is an endonuclease that makes a cut that creates the 5′ end of the mature tRNA. To produce the 3′ end of mature tRNA, an endonuclease makes a cut, and then the exonuclease RNaseD removes nine nucleotides at the 3′ end. In addition to these cleavage steps, several bases are modified to other bases, as schematically indicated. A similar type of precursor tRNA processing occurs in eukaryotes.

Concept check: What is the difference between an endonuclease and an exonuclease?

Different Splicing Mechanisms Can Remove Introns

Although the discovery of ribozymes was very surprising, the observation that tRNA and rRNA transcripts are processed to a smaller form did not seem unusual to geneticists and biochemists, because the cleavage of RNA was similar to the cleavage that can occur for other macromolecules such as DNA and proteins. In sharp contrast, when RNA splicing was detected in the 1970s, it was a novel concept. Splicing involves cleavage at two sites. An intron is removed, and—in a unique step—the remaining fragments are hooked back together again.

Since the discovery of introns, the investigations of many research groups have shown that most structural genes in complex eukaryotes contain one or more introns. Less commonly, introns can occur within tRNA and rRNA genes. At the molecular level, different RNA-splicing mechanisms have been identified. In the three examples shown in **Figure 14.18**, splicing leads to the removal of the intron RNA and the covalent connection of the exon RNA.

The splicing among **group I** and **group II introns** occurs via **self-splicing**—splicing that does not require the aid of other catalysts. Instead, the RNA functions as its own ribozyme. Group I and II differ in the ways that the introns are removed and the exons are connected. Group I introns that occur within the rRNA of *Tetrahymena* (a protozoan) have been studied extensively by Thomas Cech and colleagues. In this organism, the splicing process involves the binding of a single guanosine to a guanosine-binding site within the intron (Figure 14.18a). This guanosine breaks the bond between the first exon and the intron. The guanosine becomes attached to the 5′ end of the intron. The 3′—OH group of exon 1 then breaks the bond next to a different nucleotide (in this example, a G) that lies at the boundary between the end of the intron and exon 2; exon 1 forms a phosphoester bond with the 5′ end of exon 2. The intron RNA is subsequently degraded. In this example, the RNA molecule functions as its own ribozyme, because it splices itself without the aid of a catalytic protein.

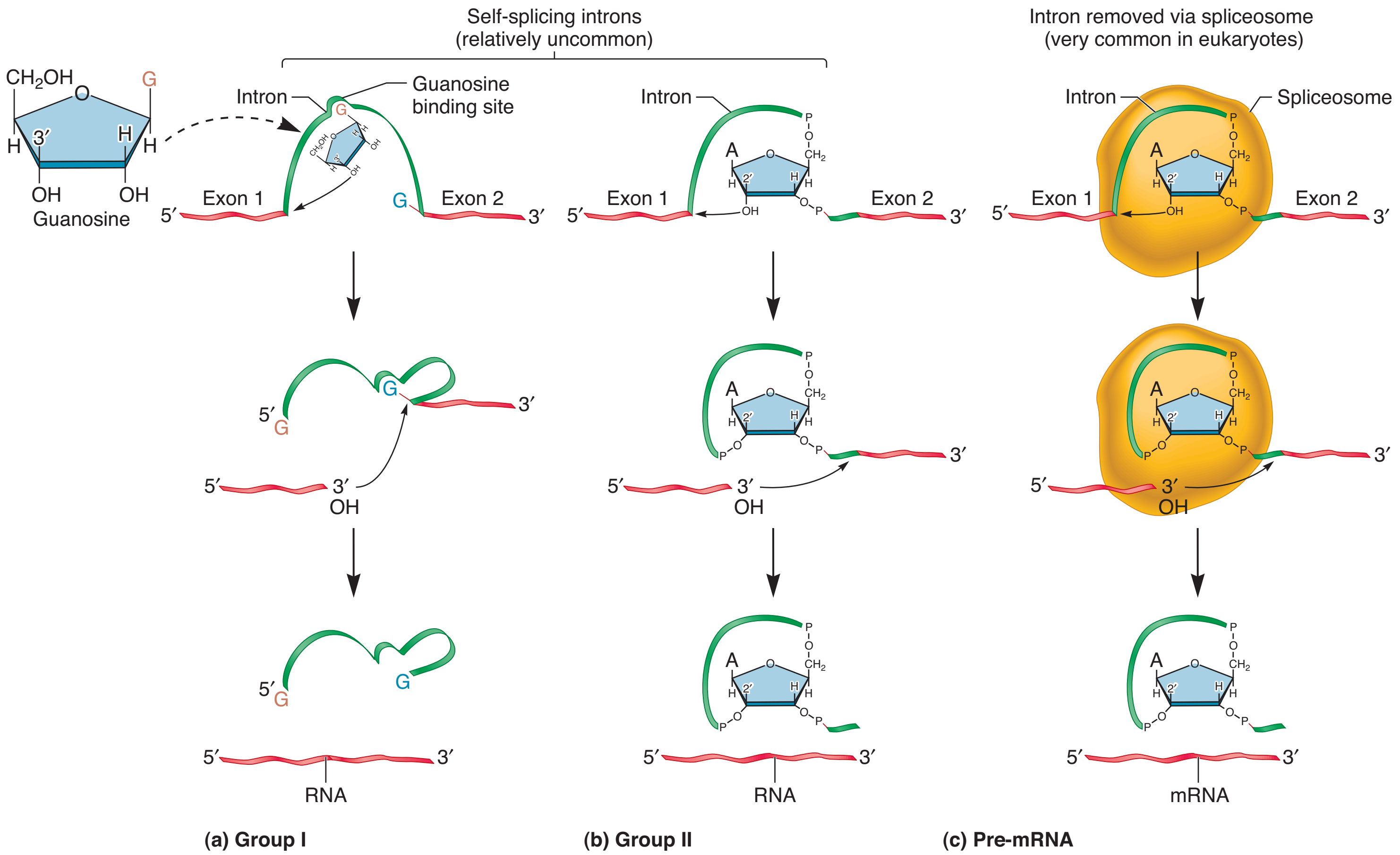
In group II introns, a similar splicing mechanism occurs, except the 2′—OH group on ribose found in an adenine nucleotide already within the intron strand begins the catalytic process (Figure 14.18b). Experimentally, group I and II self-splicing can occur in vitro without the addition of any proteins. However, in a living cell, proteins known as **maturases** often enhance the rate of splicing of group I and II introns.

In eukaryotes, the transcription of structural genes produces a long transcript known as **pre-mRNA,** which is located within the nucleus. This pre-mRNA is usually altered by splicing and other modifications before it exits the nucleus. Unlike group I and II introns, which may undergo self-splicing, pre-mRNA splicing requires the aid of a multicomponent structure known as the **spliceosome** (Figure 14.18c). As discussed shortly, this is needed to recognize the intron boundaries and to properly remove it.

Table 14.4 describes the occurrence of introns among the genes of different species. The splicing of group I and II introns is relatively uncommon. By comparison, pre-mRNA splicing is a widespread phenomenon among complex eukaryotes. In mammals and flowering plants, most structural genes have at least one intron that can be located anywhere within the gene. For example, an average human gene has about eight introns.

Pre-mRNA Splicing Occurs by the Action of a Spliceosome

The spliceosome is a large complex that splices pre-mRNA in eukaryotes. It is composed of several subunits known as **small nuclear riboproteins** (**snRNPs,** pronounced "snurps").



(a) Group I

(b) Group II

(c) Pre-mRNA

FIGURE 14.18 **Mechanisms of RNA splicing.** Group I and II introns are self-splicing. **(a)** The splicing of group I introns involves the binding of a free guanosine to a site within the intron, leading to the cleavage of RNA at the 3′ end of exon 1. The bond between a different guanine nucleotide (in the intron strand) and the 5′ end of exon 2 is cleaved. The 3′ end of exon 1 then forms a covalent bond with the 5′ end of exon 2. **(b)** In group II introns, a similar splicing mechanism occurs, except that the 2′ –OH group on an adenine nucleotide (already within the intron) begins the catalytic process. **(c)** Pre-mRNA splicing requires the aid of a multicomponent structure known as the spliceosome.

Concept check: *Which of these three mechanisms is very common in eukaryotes?*

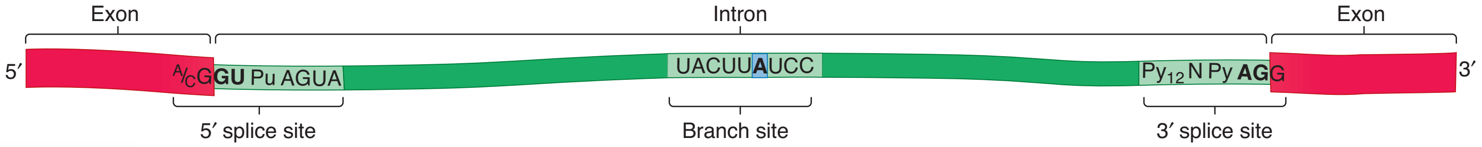
TABLE 14.4
Occurrence of Introns

Type of Intron	Mechanism of Removal	Occurrence
Group I	Self-splicing	Found in rRNA genes within the nucleus of *Tetrahymena* and other simple eukaryotes. Found in a few structural, tRNA, and rRNA genes within the mitochondrial DNA (fungi and plants) and in chloroplast DNA. Found very rarely in tRNA genes within bacteria.
Group II	Self-splicing	Found in a few structural, tRNA, and rRNA genes within the mitochondrial DNA (fungi and plants) and in chloroplast DNA. Also found rarely in bacterial genes.
Pre-mRNA	Spliceosome	Very commonly found in structural genes within the nucleus of eukaryotes.

Each snRNP contains small nuclear RNA (snRNA) and a set of proteins. During splicing, the subunits of a spliceosome carry out several functions. First, spliceosome subunits bind to an intron sequence and precisely recognize the intron-exon boundaries. In addition, the spliceosome must hold the pre-mRNA in the correct configuration to ensure the splicing together of the exons. And finally, the spliceosome catalyzes the chemical reactions that cause the introns to be removed and the exons to be covalently linked.

Intron RNA is defined by particular sequences within the intron and at the intron-exon boundaries. The consensus sequences for the splicing of mammalian pre-mRNA are shown in Figure 14.19. The bases most commonly found at these sites—those that are highly conserved evolutionarily—are shown in bold. The 5′ and 3′ splice sites occur at the ends of the intron, whereas the branch site is somewhere in the middle. These sites are recognized by components of the spliceosome.

The molecular mechanism of pre-mRNA splicing is depicted in Figure 14.20. The snRNP designated U1 binds to the 5′ splice site, and U2 binds to the branch site. This is followed by the binding of a trimer of three snRNPs: a U4/U6 dimer plus U5. The intron loops outward, and the two exons are brought closer together. The 5′ splice site is then cut, and the 5′ end of the intron becomes covalently attached to the 2′—OH group of a specific adenine nucleotide in the branch site. U1 and U4 are


Exon
Intron
Exon
5′
A/C G GU Pu AGUA
UACUUAUCC
Py_{12} N Py AG G
3′
5′ splice site
Branch site
3′ splice site

ONLINE ANIMATION

FIGURE 14.19 **Consensus sequences for pre-mRNA splicing in complex eukaryotes.** Consensus sequences exist at the intron-exon boundaries and at a branch site found within the intron itself. The adenine nucleotide shown in blue in this figure corresponds to the adenine nucleotide at the branch site in Figure 14.20. The nucleotides shown in bold are highly conserved. Designations: A/C = A or C, Pu = purine, Py = pyrimidine, N = any of the four bases.

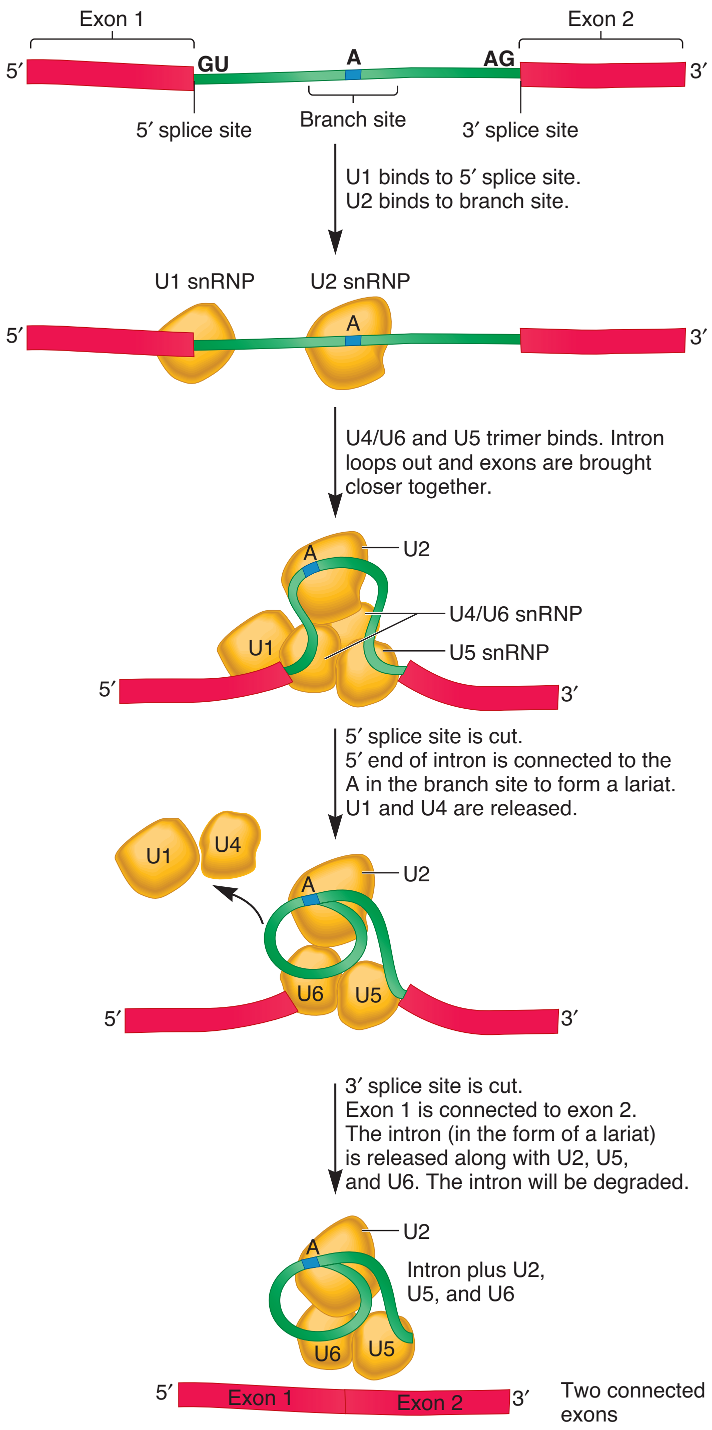
released. In the final step, the 3′ splice site is cut, and then the exons are covalently attached to each other. The three snRNPs—U2, U5, and U6—remain attached to the intron, which is in a lariat configuration. Eventually, the intron is degraded, and the snRNPs are used again to splice other pre-mRNAs.

Evidence is accumulating that certain snRNA molecules within the spliceosome play a catalytic role in the removal of introns and the connection of exons. In other words, snRNAs may function as ribozymes that cleave the RNA at the exon-intron boundaries and connect the remaining exons. Researchers have speculated that RNA molecules within U2 and U6 may have this catalytic function.

Before ending our discussion of pre-mRNA splicing, let's consider why it may be an advantage for a species to have genes that contain introns. One benefit is a phenomenon called **alternative splicing.** When a pre-mRNA has multiple introns, variation may occur in the pattern of splicing, so the resulting mRNAs contain alternative combinations of exons. The variation in splicing may happen in different cell types or during different stages of development. The biological advantage of alternative splicing is that two or more different proteins can be derived from a single gene. This allows an organism to carry fewer genes in its genome. The molecular mechanism of alternative splicing is examined in greater detail in Chapter 17. It involves the actions of proteins (not shown in Figure 14.20) that influence whether or not U1 and U2 can begin the splicing process.

The Ends of Eukaryotic Pre-mRNAs Have a 5′ Cap and a 3′ Tail

In addition to splicing, pre-mRNAs in eukaryotes are also subjected to modifications at their 5′ and 3′ ends. At their 5′ end, most mature mRNAs have a 7-methylguanosine covalently attached—an event known as **capping.** Capping occurs while the pre-mRNA is being made by RNA polymerase II, usually when the transcript is only 20 to 25 nucleotides in length. As shown in **Figure 14.21**, it is a three-step process. The nucleotide at the 5′ end of the transcript has three phosphate groups. First, an enzyme called RNA 5′-triphosphatase removes one of the phosphates, and then a second enzyme, guanylyltransferase, uses guanosine triphosphate (GTP) to attach a guanosine monophosphate (GMP) to the 5′ end. Finally, a methyltransferase attaches a methyl group to the guanine base.


Exon 1
Exon 2
5′
GU
A
AG
3′
5′ splice site
Branch site
3′ splice site
U1 binds to 5′ splice site.
U2 binds to branch site.
U1 snRNP
U2 snRNP
A
5′
3′
U4/U6 and U5 trimer binds. Intron loops out and exons are brought closer together.
A
U2
U4/U6 snRNP
U1
U5 snRNP
5′
3′
5′ splice site is cut.
5′ end of intron is connected to the A in the branch site to form a lariat.
U1 and U4 are released.
U1
U4
A
U2
U6
U5
5′
3′
3′ splice site is cut.
Exon 1 is connected to exon 2.
The intron (in the form of a lariat) is released along with U2, U5, and U6. The intron will be degraded.
A
U2
Intron plus U2, U5, and U6
U6
U5
5′
Exon 1
Exon 2
3′
Two connected exons

ONLINE ANIMATION

FIGURE 14.20 **Splicing of pre-mRNA via a spliceosome.**

Concept check: *Describe the roles of snRNPs in the splicing process.*

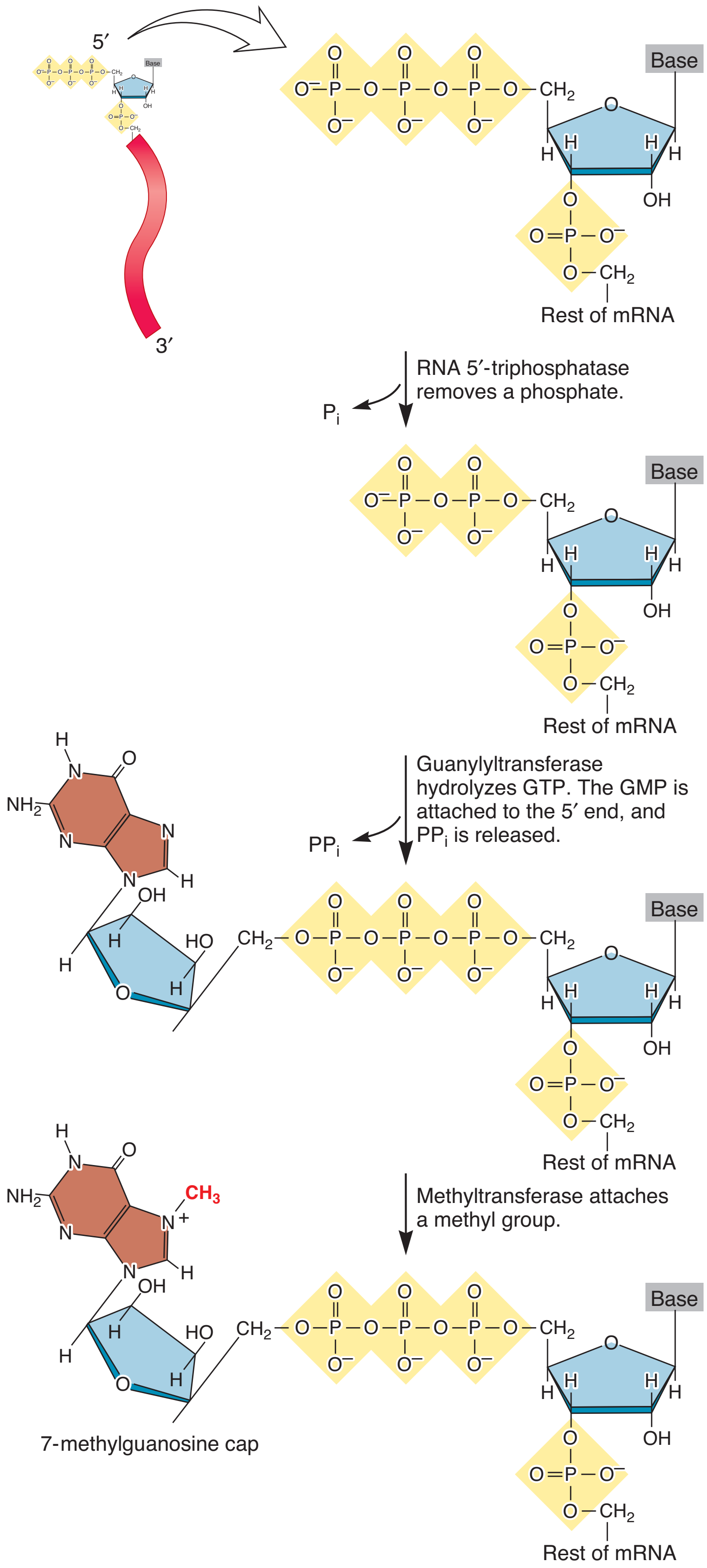

5′
3′
Base
O
CH$_2$
H
OH
O=P–O$^-$
Rest of mRNA
RNA 5′-triphosphatase removes a phosphate.
P$_i$
Guanylyltransferase hydrolyzes GTP. The GMP is attached to the 5′ end, and PP$_i$ is released.
PP$_i$
NH$_2$
N
HO
Methyltransferase attaches a methyl group.
CH$_3$
N+
7-methylguanosine cap

FIGURE 14.21 **Attachment of a 7-methylguanosine cap to the 5′ end of mRNA.** When the transcript is about 20 to 25 nucleotides in length, RNA 5′-triphosphatase removes one of the three phosphates, and then a second enzyme, guanylyltransferase, attaches GMP to the 5′ end. Finally, a methyltransferase attaches a methyl group to the guanine base.

Concept check: What are three functional roles of the 7-methlyguanosine cap?

What are the functions of the 7-methylguanosine cap? The cap structure is recognized by cap-binding proteins, which perform various roles. For example, cap-binding proteins are required for the proper exit of most mRNAs from the nucleus. Also, the cap structure is recognized by initiation factors that are needed during the early stages of translation. Finally, the cap structure may be important in the efficient splicing of introns, particularly the first intron located nearest the 5′ end.

Let's now turn our attention to the 3′ end of the RNA molecule. Most mature mRNAs have a string of adenine nucleotides, referred to as a **polyA tail,** which is important for mRNA stability, the exit of mRNA from the nucleus, and in the synthesis of polypeptides. The polyA tail is not encoded in the gene sequence. Instead, it is added enzymatically after the pre-mRNA has been completely transcribed—a process termed **polyadenylation.**

The steps required to synthesize a polyA tail are shown in Figure 14.22. To acquire a polyA tail, the pre-mRNA contains a polyadenylation signal sequence near its 3′ end. In mammals, the consensus sequence is AAUAAA. This sequence is downstream (toward the 3′ end) from the stop codon in the pre-mRNA. An endonuclease recognizes the signal sequence and cleaves the pre-mRNA at a location that is about 20 nucleotides beyond the 3′ end of the AAUAAA sequence. The fragment beyond the 3′ cut is degraded. Next, an enzyme known as polyA-polymerase attaches many adenine-containing nucleotides. The length of the polyA tail varies among different mRNAs, from a few dozen to several hundred adenine nucleotides. As discussed in Chapter 17, a long polyA tail increases the stability of mRNA.

The Nucleotide Sequence of RNA Can Be Modified by RNA Editing

The term **RNA editing** refers to a change in the nucleotide sequence of an RNA molecule that involves additions or deletions of particular bases, or a conversion of one type of base to another, such as a cytosine to a uracil. In the case of mRNAs, editing can have various effects, such as generating start codons, generating stop codons, and changing the coding sequence for a polypeptide.

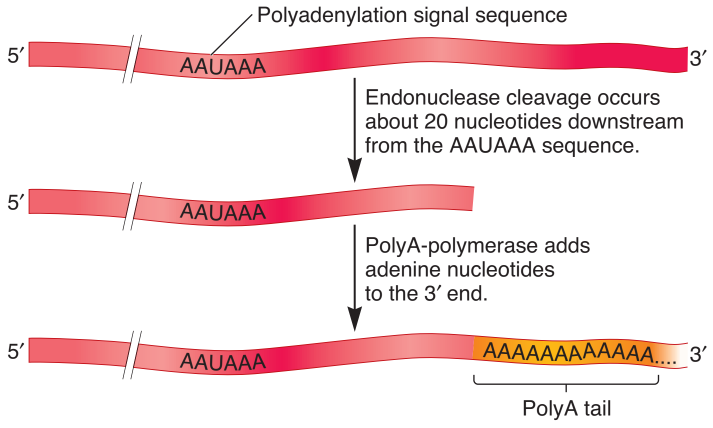

Polyadenylation signal sequence
5′
AAUAAA
3′
Endonuclease cleavage occurs about 20 nucleotides downstream from the AAUAAA sequence.
5′
AAUAAA
PolyA-polymerase adds adenine nucleotides to the 3′ end.
5′
AAUAAA
AAAAAAAAAAAA....
3′
PolyA tail

FIGURE 14.22 **Attachment of a polyA tail.** First, an endonuclease cuts the RNA at a location that is 11 to 30 nucleotides after the AAUAAA polyadenylation sequence, making the RNA shorter at its 3′ end. Adenine-containing nucleotides are then attached, one at a time, to the 3′ end by the enzyme polyA-polymerase.

The phenomenon of RNA editing was first discovered in trypanosomes, the protists that cause sleeping sickness. Since that time, however, RNA editing has been shown to occur in various organisms and in a variety of ways, although its functional significance is slowly emerging (**Table 14.5**). In the specific case of trypanosomes, the editing process involves the addition or deletion of one or more uracil nucleotides in an RNA.

A more common mechanism for RNA editing involves changes of one type of base to another. In this form of editing, a base in the RNA is deaminated—an amino group is removed from the base. When cytosine is deaminated, uracil is formed, and when adenine is deaminated, inosine (I) is formed (**Figure 14.23**). Inosine is recognized as guanine during translation.

An example of RNA editing occurs in mammals involving an mRNA that encodes a protein called apolipoprotein B. In the liver, translation of an unedited mRNA produces apolipoprotein B-100, a protein that is essential for the transport of cholesterol in the blood. In intestinal cells, the mRNA may be edited so that a single C is changed to a U. What is the significance of this base substitution? This change converts a glutamine codon (CAA) to a stop codon (UAA), thereby resulting in a shorter apolipoprotein. In this case, RNA editing produces an apolipoprotein B with an altered structure. Therefore, RNA editing can produce two proteins from the same gene, much like the phenomenon of alternative splicing.

How widespread is RNA editing that involves C to U and A to I substitutions? In invertebrates such as *Drosophila*, researchers estimate that 50 to 100 RNAs are edited for the purpose of changing the RNA coding sequence. In mammals, the RNAs from less than 25 genes are currently known to be edited.

14.4 REVIEWING THE KEY CONCEPTS

- RNA transcripts can be modified in a variety of ways, which include processing, splicing, 5′ capping, 3′ polyA tailing, RNA editing, and base modification (see Table 14.3).
- Certain RNA molecules such as rRNAs and tRNAs are processed to smaller, functional molecules via cleavage steps (see Figures 14.16, 14.17).
- Group I and II introns are removed by self-splicing. Pre-mRNA introns are removed via a spliceosome (see Table 14.4, Figure 14.18).
- The spliceosome is a multicomponent structure that recognizes intron sequences and removes them from pre-mRNA (see Figures 14.19, 14.20).
- In eukaryotes, mRNA is given a 7-methylguanosine cap at the 5′ end and polyA tail at the 3′ end (see Figures 14.21, 14.22).

TABLE **14.5**

Examples of RNA Editing

Organism	Type of Editing	Found In
Trypanosomes (protozoa)	Primarily additions but occasionally deletions of uracil nucleotides	Many mitochondrial mRNAs
Land plants	C-to-U conversion	Many mitochondrial and chloroplast mRNAs, tRNAs, and rRNAs
Mammals	C-to-U conversion	Apolipoprotein B mRNA
	A-to-I conversion	Glutamate receptor mRNA, many tRNAs
Drosophila	A-to-I conversion	mRNA for calcium and sodium channels

I, inosine

- RNA editing changes the base sequence of an RNA after it has been synthesized (see Table 14.5, Figure 14.23).

14.4 COMPREHENSION QUESTIONS

1. Which of the following are examples of RNA modification?
   a. Splicing
   b. 7-methylguanosine cap
   c. PolyA tail
   d. All of the above
2. A ribozyme is
   a. a complex between RNA and a protein.
   b. an RNA than encodes a protein that functions as an enzyme.
   c. an RNA molecule with catalytic function.
   d. a protein that degrades RNA molecules.
3. Which of the following statements about the spliceosome is *false*?
   a. A spliceosome splices pre-mRNA molecules.
   b. A spliceosome removes exons from RNA molecules.
   c. A spliceosome is composed of snRNPs.
   d. A spliceosome recognizes the exon-intron boundaries and the branch site.
4. Which of the following is a function of the 7-methylguanosine cap?
   a. Exit of mRNA from the nucleus
   b. Efficient splicing of pre-mRNA
   c. Initiation of translation
   d. All of the above

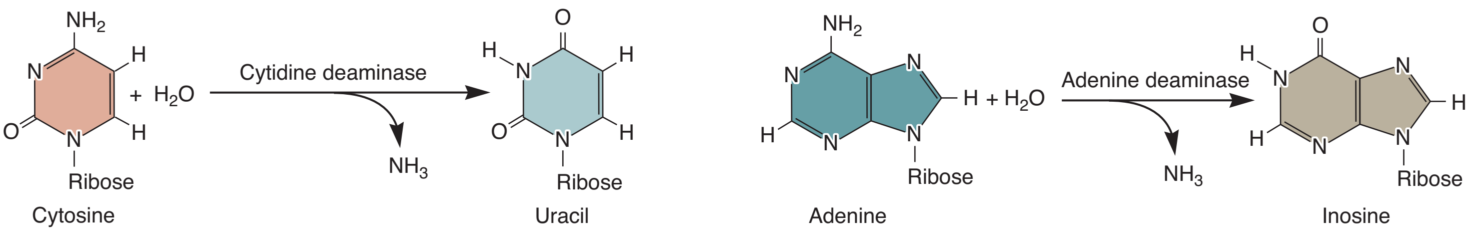


FIGURE 14.23 **RNA editing by deamination.** A cytidine deaminase can remove an amino group from cytosine, thereby creating uracil. An adenine deaminase can remove an amino group from adenine to make inosine.

Concept check: *What is a functional consequence of RNA editing?*

KEY TERMS

Page 299. gene, transcription, structural genes, messenger RNA (mRNA), translation, central dogma of genetics, gene expression

Page 301. promoter, terminator, transcription factors, regulatory sequences, regulatory elements, ribosome-binding site, codons, start codon, stop codon, initiation, elongation, termination, RNA polymerase

Page 302. open complex

Page 303. transcriptional start site, sequence elements, Pribnow box, consensus sequence

Page 304. core enzyme, sigma (σ) factor, holoenzyme, helix-turn-helix motif, closed complex

Page 305. template strand, coding strand

Page 306. (ρ) rho, ρ-dependent termination, ρ-independent termination, intrinsic termination

Page 308. core promoter, TATA box, basal transcription, enhancers

Page 309. silencers, *cis*-acting elements, *trans*-acting factors, general transcription factors (GTFs)

Page 310. preinitiation complex, basal transcription apparatus, mediator

Page 312. colinearity, exons, intervening sequences, introns, RNA splicing, nucleolus, exonuclease, endonuclease

Page 313. ribozyme

Page 314. group I introns, group II introns, self-splicing, maturases, pre-mRNA, spliceosome, small nuclear riboproteins (snRNPs)

Page 316. alternative splicing, capping

Page 317. polyA tail, polyadenylation, RNA editing

CHAPTER SUMMARY

- According to the central dogma of genetics, DNA is transcribed into mRNA, and mRNA is translated into a polypeptide. DNA replication allows the DNA to be passed from cell to cell and from parent to offspring (see Figure 14.1).

14.1 Overview of Transcription

- A gene is an organization of DNA sequences. A promoter signals the start of transcription, and a terminator signals the end. Regulatory sequences control the rate of transcription. For genes that encode polypeptides, the gene sequence also specifies a start codon, a stop codon, and many codons in between. Bacterial genes also specify a ribosomal binding sequence (see Figure 14.2).
- Transcription occurs in three phases called initiation, elongation, and termination (see Figure 14.3).
- RNA transcripts have many different functions (see Table 14.1).

14.2 Transcription in Bacteria

- Many bacterial promoters have sequence elements at the −35 and −10 regions. The transcriptional start site is at +1 (see Figures 14.4, 14.5).
- During the initiation phase of transcription in *E. coli*, sigma (σ) factor, which is bound to RNA polymerase, binds into the major groove of DNA and recognizes sequence elements at the promoter. This process forms a closed complex. Following the formation of an open complex, σ factor is released (see Figures 14.6, 14.7).
- During the elongation phase of transcription, RNA polymerase slides along the DNA and maintains an open complex as it goes. RNA is made in the 5′ to 3′ direction according to complementary base pairing (see Figure 14.8).
- In a given chromosome, the use of the template strand varies from gene to gene (see Figure 14.9).
- Transcriptional termination in *E. coli* occurs by a rho (ρ)-dependent or ρ-independent mechanism (see Figures 14.10, 14.11).

14.3 Transcription in Eukaryotes

- Eukaryotes use RNA polymerase I, II, and III to transcribe different categories of genes. Prokaryotic and eukaryotic RNA polymerases have similar structures (see Figure 14.12).
- Eukaryotic promoters have a core promoter and regulatory elements such as enhancers and silencers (see Figure 14.13).
- Transcription of structural genes in eukaryotes requires RNA polymerase II, five general transcription factors, and mediator. The five general transcription factors and RNA polymerase assemble together to form an open complex (see Table 14.2, Figure 14.14).
- Transcriptional termination of RNA polymerase II may occur via the allosteric or torpedo model (see Figure 14.15).

14.4 RNA Modification

- RNA transcripts can be modified in a variety of ways, which include processing, splicing, 5′ capping, 3′ polyA tailing, RNA editing, and base modification (see Table 14.3).
- Certain RNA molecules such as ribosomal RNAs and precursor tRNAs are processed to smaller, functional molecules via cleavage steps (see Figures 14.16, 14.17).
- Group I and II introns are removed by self-splicing. Pre-mRNA introns are removed via a spliceosome (see Table 14.4, Figure 14.18).
- The spliceosome is a multicomponent structure that recognizes intron sequences and removes them from pre-mRNA (see Figures 14.19, 14.20).
- In eukaryotes, mRNA is given a 7-methylguanosine cap at the 5′ end and polyA tail at the 3′ end (see Figures 14.21, 14.22).
- RNA editing changes the base sequence of an RNA after it has been synthesized (see Table 14.5, Figure 14.23).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe the important events that occur during the three stages of gene transcription in bacteria. What proteins play critical roles in the three stages?

Answer: The three stages are initiation, elongation, and termination.

Initiation: RNA polymerase holoenzyme slides along the DNA until σ factor recognizes a promoter. σ factor binds tightly to this sequence, forming a closed complex. The DNA is then denatured to form a bubble-like structure known as the open complex.

Elongation: RNA polymerase core enzyme slides along the DNA, synthesizing RNA as it goes. The α subunits of RNA polymerase keep the enzyme bound to the DNA, while the β subunits are responsible for binding and for the catalytic synthesis of RNA. The ϖ (omega) subunit is also important for the proper assembly of the core enzyme. During elongation, RNA is made according to the AU/GC rule, with nucleotides being added in the 5′ to 3′ direction.

Termination: RNA polymerase eventually reaches a sequence at the end of the gene that signals the end of transcription. In ρ-independent termination, the properties of the termination sequences in the DNA are sufficient to cause termination. In ρ-dependent termination, the ρ (rho) protein recognizes a sequence within the RNA, binds there, and travels toward RNA polymerase. When the formation of a stem-loop causes RNA polymerase to pause, ρ catches up and separates the RNA-DNA hybrid region, releasing RNA polymerase.

S2. What is the difference between a structural gene and a nonstructural gene?

Answer: Structural genes encode mRNAs that are translated into polypeptide sequences. Nonstructural genes encode RNAs that are never translated. Products of nonstructural genes include tRNA and rRNA, which function during translation; microRNA, which is involved in gene regulation; 7S RNA, which is part of a complex known as SRP; scRNA, small cytoplasmic RNA found in bacteria; the RNA of RNaseP; telomerase RNA, which is involved in telomere replication; snoRNA, which is involved in rRNA trimming; and snRNA, which is a component of spliceosome. In many cases, the RNA from nonstructural genes becomes part of a complex composed of RNA molecules and protein subunits.

S3. When RNA polymerase transcribes DNA, only one of the two DNA strands is used as a template. Take a look at Figure 14.4 and explain how RNA polymerase determines which DNA strand is the template strand.

Answer: The binding of σ factor and RNA polymerase depends on the sequence of the promoter. RNA polymerase binds to the promoter in such a way that the –35 sequence TTGACA and the –10 sequence TATAAT are within the coding strand, whereas the –35 sequence AACTGT and the –10 sequence ATATTA are within the template strand.

Conceptual Questions

C1. Genes may be structural genes that encode polypeptides, or they may be nonstructural genes.

A. Describe three examples of genes that are not structural genes.

B. For structural genes, one DNA strand is called the template strand, and the complementary strand is called the coding strand. Are these two terms appropriate for nonstructural genes? Explain.

C. Do nonstructural genes have a promoter and a terminator?

C2. In bacteria, what event marks the end of the initiation stage of transcription?

C3. What is the meaning of the term *consensus sequence*? Give an example. Describe the locations of consensus sequences within bacterial promoters. What are their functions?

C4. What is the consensus sequence of the following six DNA sequences?

GGCATTGACT
GCCATTGTCA
CGCATAGTCA
GGAAATGGGA
GGCTTTGTCA
GGCATAGTCA

C5. Mutations in bacterial promoters may increase or decrease gene transcription. Promoter mutations that increase transcription are termed *up-promoter* mutations, and those that decrease transcription are termed *down-promoter* mutations. The sequence of the −10 region of the promoter for the *lac* operon is TATGTT (see Figure 14.5). Would you expect the following mutations to be up-promoter or down-promoter mutations?

A. TATGTT to TATATT

B. TATGTT to TTTGTT

C. TATGTT to TATGAT

C6. In the examples shown in Figure 14.5, which positions of the −35 sequence (i.e., first, second, third, fourth, fifth, or sixth) are more tolerant of changes? Do you think these positions play a more or less important role in the binding of σ factor? Explain why.

C7. In Chapter 11, we considered the dimensions of the double helix (see Figure 11.14). In an α helix of a protein, there are 3.6 amino acids per complete turn. Each amino acid advances the α helix by 0.15 nm; a complete turn of an α helix is 0.54 nm in length. As shown in Figure 14.6, two α helices of a transcription factor occupy the major groove of the DNA. According to Figure 14.6, estimate the number of amino acids that bind to this region. How many complete turns of the α helices occupy the major groove of DNA?

C8. A mutation within a gene changes the start codon to a stop codon. How will this mutation affect the transcription of this gene?

C9. What is the subunit composition of bacterial RNA polymerase holoenzyme? What are the functional roles of the subunits?

C10. At the molecular level, describe how σ factor recognizes bacterial promoters. Be specific about the structure of σ factor and the type of chemical bonding.

C11. Let's suppose a mutation changes the consensus sequence at the –35 location in a way that inhibits σ factor binding. Explain how a mutation would inhibit σ factor from binding to the DNA. Look at Figure 14.5 and describe two specific base substitutions you think would inhibit the binding of σ factor. Explain why your base substitutions would have this effect.

C12. What is the complementarity rule that governs the synthesis of an RNA molecule during transcription? An RNA transcript has the following sequence:

5′–GGCAUGCAUUACGGCAUCACACUAGGGAUC–3′

What is the sequence of the template and coding strands of the DNA that encodes this RNA? On which side (5′ or 3′) of the template strand is the promoter located?

C13. Describe the movement of the open complex along the DNA.

C14. Describe what happens to the chemical bonding interactions when transcriptional termination occurs. Be specific about the type of chemical bonding.

C15. Discuss the differences between ρ-dependent and ρ-independent termination.

C16. In Chapter 13, we discussed the function of DNA helicase, which is involved in DNA replication. Discuss how the functions of helicase and ρ protein are similar and how they are different.

C17. Discuss the similarities and differences between RNA polymerase and DNA polymerase (described in Chapter 13).

C18. Mutations that occur at the end of a gene may alter the sequence of the gene and prevent transcriptional termination.

A. What types of mutations would prevent ρ-independent termination?

B. What types of mutations would prevent ρ-dependent termination?

C. If a mutation prevented transcriptional termination at the end of a gene, where would gene transcription end? Or would it end?

C19. If the following RNA polymerases were missing from a eukaryotic cell, what types of genes would not be transcribed?

A. RNA polymerase I

B. RNA polymerase II

C. RNA polymerase III

C20. What sequence elements are found within the core promoter of structural genes in eukaryotes? Describe their locations and specific functions.

C21. For each of the following transcription factors, how would eukaryotic transcriptional initiation be affected if it were missing?

A. TFIIB

B. TFIID

C. TFIIH

C22. Describe the allosteric and torpedo models for transcriptional termination of RNA polymerase II. Which model is more similar to ρ-dependent termination in bacteria and which model is more similar to ρ-independent termination?

C23. Which eukaryotic transcription factor(s) shown in Figure 14.14 plays an equivalent role to σ factor found in bacterial cells?

C24. The initiation phase of eukaryotic transcription via RNA polymerase II is considered an assembly and disassembly process. Which types of biochemical interactions—hydrogen bonding, ionic bonding, covalent bonding, and/or hydrophobic interactions—would you expect to drive the assembly and disassembly process? How would temperature and salt concentration affect assembly and disassembly?

C25. A eukaryotic structural gene contains two introns and three exons: exon 1–intron 1–exon 2–intron 2–exon 3. The 5′ splice site at the boundary between exon 2 and intron 2 has been eliminated by a small deletion in the gene. Describe how the pre-mRNA encoded by this mutant gene would be spliced. Indicate which introns and exons would be found in the mRNA after splicing occurs.

C26. Describe the processing events that occur during the production of tRNA in *E. coli*.

C27. Describe the structure and function of a spliceosome. Speculate why the spliceosome subunits contain snRNA. In other words, what do you think is/are the functional role(s) of snRNA during splicing?

C28. What is the unique feature of ribozyme function? Give two examples described in this chapter.

C29. What does it mean to say that gene expression is colinear?

C30. What is meant by the term *self-splicing*? What types of introns are self-splicing?

C31. In eukaryotes, what types of modifications occur to pre-mRNAs?

C32. What is alternative splicing? What is its biological significance?

C33. The processing of ribosomal RNA in eukaryotes is shown in Figure 14.16. Why is this called cleavage or processing but not splicing?

C34. In the splicing of group I introns shown in Figure 14.18, does the 5′ end of the intron have a phosphate group? Explain.

C35. According to the mechanism shown in Figure 14.20, several snRNPs play different roles in the splicing of pre-mRNA. Identify the snRNP that recognizes the following sites:

A. 5′ splice site

B. 3′ splice site

C. Branch site

C36. After the intron (which is in a lariat configuration) is released during pre-mRNA splicing, a brief moment occurs before the two exons are connected to each other. Which snRNP(s) holds the exons in place so they can be covalently connected to each other?

C37. A lariat contains a closed loop and a linear end. An intron has the following sequence: 5′–GUPuAGUA–60 nucleotides–UACUUAUCC–100 nucleotides–Py_{12}NPyAG–3′. Which sequence would be found within the closed loop of the lariat, the 60-nucleotide sequence or the 100-nucleotide sequence?

Application and Experimental Questions

E1. A research group has sequenced the cDNA and genomic DNA from a particular gene. The cDNA is derived from mRNA, so it does not contain introns. Here are the DNA sequences.

cDNA:

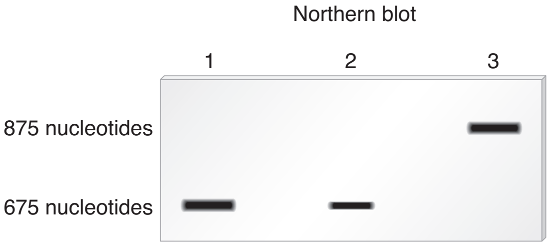
5′–ATTGCATCCAGCGTATACTATCTCGGGCCCAATTAAT -
GCCAGCGGCCAGACTATCACCCAACTCGGTTACCTACTAG -
TATATCCCATATACTAGCATATATTTTACCCATAATTTGTGTGT -
GGGTATACAGTATAATCATATA–3′

Genomic DNA (contains one intron):

5′-ATTGCATCCAGCGTATACTATCTCGGGCCCAAT -
TAATGCCAGCGGCCAGACTATCACCCAACTCG -
GCCCACCCCCCAGGTTTACACAGTCATACCATACA
TACAAAAATCGCAGTTACTTATCCCAAAAAAACCTAG -
ATACCCCACATACTATTAACTCTTTCTTTCTAGGTTACCTAC -
TAGTATATCCCATATACTAGCATATATTTTACCCATAATTTGT -
GTGTGGGTATACAGTATAATCATATA–3′

Indicate where the intron is located. Does the intron contain the normal consensus splice site sequences based on those described in Figure 14.19? Underline the splice site sequences, and indicate whether or not they fit the consensus sequence.

E2. Chapter 19 describes a technique known as Northern blotting that can be used to detect RNA transcribed from a particular gene. In this method, a specific RNA is detected using a short segment of cloned DNA as a probe. The DNA probe, which is radioactive, is complementary to the RNA that the researcher wishes to detect. After the radioactive probe DNA binds to the RNA, the RNA is visualized as a dark (radioactive) band on an X-ray film. As shown here, the method of Northern blotting can be used to determine the amount of a particular RNA transcribed in a given cell type. If one type of cell produces twice as much of a particular mRNA as occurs in another cell, the band will appear twice as intense. Also, the method can distinguish if alternative RNA splicing has occurred to produce an RNA that has a different molecular mass.



Lane 1 is a sample of RNA isolated from nerve cells.

Lane 2 is a sample of RNA isolated from kidney cells. Nerve cells produce twice as much of this RNA as do kidney cells.

Lane 3 is a sample of RNA isolated from spleen cells. Spleen cells produce an alternatively spliced version of this RNA that is about 200 nucleotides longer than the RNA produced in nerve and kidney cells.

Let's suppose a researcher was interested in the effects of mutations on the expression of a particular structural gene in eukaryotes. The gene has one intron that is 450 nucleotides long. After this intron is removed from the pre-mRNA, the mRNA transcript is 1100 nucleotides in length. Diploid somatic cells have two copies of this gene. Make a drawing that shows the expected results of a Northern blot using mRNA from the cytosol of somatic cells, which were obtained from the following individuals:

Lane 1: A normal individual

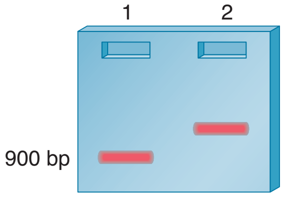
Lane 2: A homozygote for a deletion that removes the −50 to −100 region of the gene that encodes this mRNA

Lane 3: A heterozygote in which one gene is normal and the other gene had a deletion that removes the −50 to −100 region

Lane 4: A homozygote for a mutation that introduces an early stop codon into the middle of the coding sequence of the gene

Lane 5: A homozygote for a three-nucleotide deletion that removes the AG sequence at the 3′ splice site

E3. A gel retardation assay can be used to study the binding of proteins to a segment of DNA. This method is described in Chapter 19. When a protein binds to a segment of DNA, it retards the movement of the DNA through a gel, so the DNA appears at a higher point in the gel (see the following).



Lane 1: 900-bp fragment alone

Lane 2: 900-bp fragment plus a protein that binds to the 900-bp fragment

In this example, the segment of DNA is 900 bp in length, and the binding of a protein causes the DNA to appear at a higher point in the gel. If this 900-bp fragment of DNA contains a eukaryotic promoter for a structural gene, draw a gel that shows the relative locations of the 900-bp fragment under the following conditions:

Lane 1: 900 bp plus TFIID

Lane 2: 900 bp plus TFIIB

Lane 3: 900 bp plus TFIID and TFIIB

Lane 4: 900 bp plus TFIIB and RNA polymerase II

Lane 5: 900 bp plus TFIID, TFIIB, and RNA polymerase II/TFIIF

E4. As described in Chapter 19 and in experimental question E3, a gel retardation assay can be used to determine if a protein binds to DNA. This method can also determine if a protein binds to RNA. In the combinations described here, would you expect the migration of the RNA to be retarded due to the binding of a protein?

A. mRNA from a gene that is terminated in a ρ-independent manner plus ρ-protein

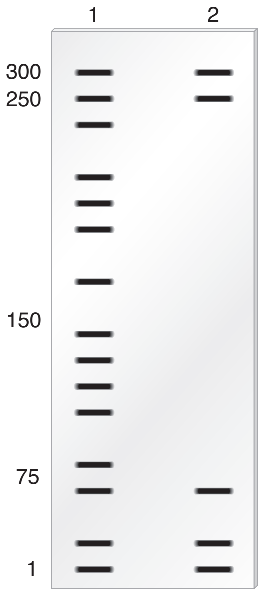
B. mRNA from a gene that is terminated in a ρ-dependent manner plus ρ-protein

C. pre-mRNA from a structural gene that contains two introns plus the snRNP called U1

D. Mature mRNA from a structural gene that contains two introns plus the snRNP called U1

E5. The technique of DNase I footprinting is described in Chapter 19. If a protein binds over a region of DNA, it will protect chromatin in that region from digestion by DNase I. To carry out a DNA footprinting experiment, a researcher has a sample of a cloned DNA fragment. The fragments are exposed to DNase I in the presence and absence of a DNA-binding protein. Regions of the DNA fragment not covered by the DNA-binding protein are digested by DNase I, producing a series of bands on a gel. Regions of the DNA fragment not digested by DNase I (because a DNA-binding protein is preventing DNase I from gaining access to the DNA) are revealed, because a region of the gel does not contain any bands.

In the DNase I footprinting experiment shown here, a researcher began with a sample of cloned DNA 300 bp in length. This DNA contained a eukaryotic promoter for RNA polymerase II. For the sample loaded in lane 1, no proteins were added. For the sample loaded in lane 2, the 300-bp fragment was mixed with RNA polymerase II plus TFIID and TFIIB.



A. How long of a region of DNA is "covered up" by the binding of RNA polymerase II and the transcription factors?

B. Describe how this binding would occur if the DNA was within a nucleosome structure. (Note: The structure of nucleosomes is described in Chapter 12.) Do you think that the DNA is in a nucleosome structure when RNA polymerase and transcription factors are bound to the promoter? Explain why or why not.

E6. As described in Table 14.1, several different types of RNA are made, especially in eukaryotic cells. Researchers are sometimes interested in focusing their attention on the transcription of structural genes in eukaryotes. Such researchers want to study mRNA. One method that is used to isolate mRNA is column chromatography. (Note: See the Appendix for a general description of chromatography.) Researchers can covalently attach short pieces of DNA that contain stretches of thymine (i.e., TTTTTTTTTTTT) to the column matrix. This is called a poly-dT column. When a cell extract is poured over the column, mRNA binds to the column, but other types of RNA do not.

A. Explain how you would use a poly-dT column to obtain a purified preparation of mRNA from eukaryotic cells. In your description, explain why mRNA binds to this column and what you would do to release the mRNA from the column.

B. Can you think of ways to purify other types of RNA, such as tRNA or rRNA?

Questions for Student Discussion/Collaboration

1. Based on your knowledge of introns and pre-mRNA splicing, discuss whether or not you think alternative splicing fully explains the existence of introns. Can you think of other possible reasons to explain the existence of introns?

2. Discuss the types of RNA transcripts and the functional roles they play. Why do some RNAs form complexes with protein subunits?

Answers to Comprehension Questions

14.1: a, c, a

14.2: d, b, a, b

14.3: b, c, d, b

14.4: d, c, b, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

15

CHAPTER OUTLINE

15.1 The Genetic Basis for Protein Synthesis

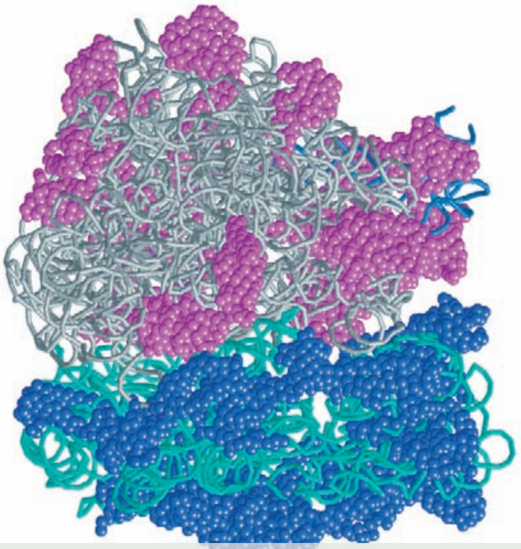
15.2 The Relationship Between the Genetic Code and Protein Synthesis

15.3 Experimental Determination of the Genetic Code

15.4 Structure and Function of tRNA

15.5 Ribosome Structure and Assembly

15.6 Stages of Translation



This is a model of ribosome structure based on X-ray crystallography. *Ribosomes are needed to synthesize polypeptides, using mRNA as a template. A detailed description of this model is discussed later in Figure 15.14.*

TRANSLATION OF mRNA

In Chapter 14, we considered transcription in which the DNA sequence within a gene is copied into a nucleotide sequence of messenger RNA (mRNA). During **translation,** mRNA is used to specify the sequence of amino acids within a polypeptide (refer back to Figure 14.1). In this chapter, we will explore the current state of knowledge regarding translation, focusing on the specific molecular interactions responsible for this process. During the past few decades, the concerted efforts of geneticists, cell biologists, and biochemists have profoundly advanced our understanding of translation. Even so, many questions remain unanswered, and this topic continues to be an exciting area of investigation.

We will begin by considering the classic experiments that revealed the purpose of some genes is to encode proteins that function as enzymes. Next, we examine how the genetic code is used to decipher the information within mRNA to produce a polypeptide with a specific amino acid sequence. The rest of this chapter is devoted to an understanding of translation at the molecular level as it occurs in living cells. This will involve an examination of the cellular components—including many different proteins, RNAs, and small molecules—needed for the translation process. We will consider the structure and function of tRNA molecules, which act as the translators of the genetic information within mRNA, and then examine the composition of ribosomes. Finally, we will explore the three stages of translation in bacterial cells and eukaryotic cells.

15.1 THE GENETIC BASIS FOR PROTEIN SYNTHESIS

Proteins are critically important as active participants in cell structure and function. The primary role of DNA is to store the information needed for the synthesis of all the proteins that an organism makes. As we discussed in Chapter 14, genes that encode an amino acid sequence are known as **structural genes.** The RNA transcribed from structural genes is called **messenger RNA (mRNA).** The main function of the genetic material is to encode the production of cellular proteins in the correct cell, at the proper time, and in suitable amounts. This is an extremely complicated task because living cells make thousands of different proteins. Genetic analyses have shown that a typical bacterium can make a few thousand different proteins, and estimates for eukaryotes range from several thousand in simple eukaryotic organisms, such as yeast, to tens of thousands in plants and animals. In this section, we will consider early experiments which showed that the role of some genes is to encode enzymes.

Archibald Garrod Proposed That Some Genes Code for the Production of a Single Enzyme

The idea that a relationship exists between genes and the production of proteins was first suggested at the beginning of the twentieth century by Archibald Garrod, a British physician. Prior to Garrod's studies, biochemists had studied many metabolic pathways within living cells. These pathways consist of a series of metabolic conversions of one molecule to another, each step catalyzed by a specific enzyme. Each enzyme is a distinctly different protein that catalyzes a particular chemical reaction. **Figure 15.1** illustrates part of the metabolic pathway for the degradation of phenylalanine, an amino acid commonly found in human diets. The enzyme phenylalanine hydroxylase catalyzes the conversion of phenylalanine to tyrosine, and a different enzyme, tyrosine aminotransferase, converts tyrosine into *p*-hydroxyphenylpyruvic acid, and so on. In all of the steps shown in Figure 15.1, a specific enzyme catalyzes a single type of chemical reaction.

Garrod studied patients who had defects in their ability to metabolize certain compounds. He was particularly interested in the inherited disease known as **alkaptonuria.** In this disorder, the patient's body accumulates abnormal levels of homogentisic acid (also called alkapton), which is excreted in the urine, causing it to appear black on exposure to air. In addition, the disease is characterized by bluish black discoloration of cartilage and skin (ochronosis). Garrod proposed that the accumulation of homogentisic acid in these patients is due to a missing enzyme, namely, homogentisic acid oxidase (see Figure 15.1).

How did Garrod realize that certain genes encode enzymes? He already knew that alkaptonuria is an inherited trait that follows an autosomal recessive pattern of inheritance. Therefore, an individual with alkaptonuria must have inherited the mutant (defective) gene that causes this disorder from both parents. From these observations, Garrod proposed that a relationship exists between the inheritance of the trait and the inheritance of a defective enzyme. Namely, if an individual inherited the mutant gene (which causes a loss of enzyme function), she or he would not produce any normal enzyme and would be unable to metabolize homogentisic acid. Garrod described alkaptonuria as an **inborn error of metabolism.** This hypothesis was the first suggestion that a connection exists between the function of genes and the production of enzymes. At the turn of the century, this idea was particularly insightful, because the structure and function of the genetic material were completely unknown.

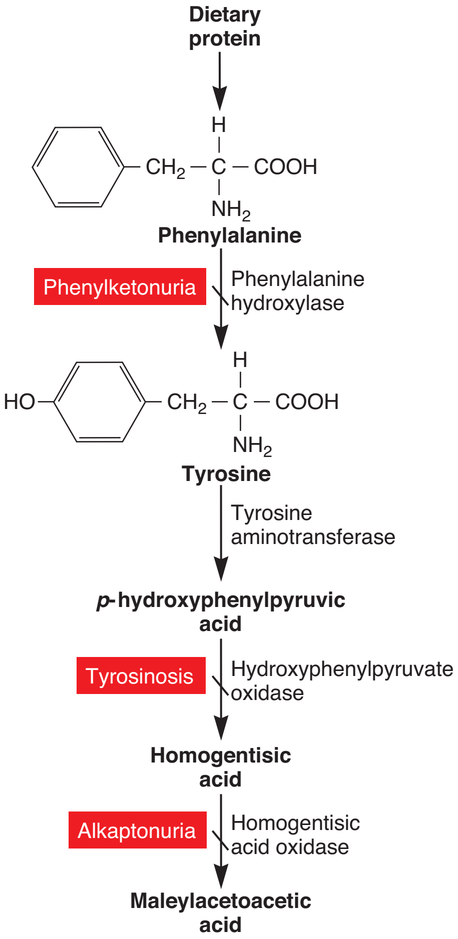


FIGURE 15.1 **The metabolic pathway of phenylalanine breakdown.** This diagram shows part of the pathway of phenylalanine metabolism, which consists of enzymes that successively convert one molecule to another. Certain human genetic diseases (shown in red boxes) are caused when enzymes in this pathway are missing or defective (indicated by a slash).

Genes → Traits When a person inherits two defective copies of the gene that encodes homogentisic acid oxidase, he or she cannot convert homogentisic acid into maleylacetoacetic acid. Such a person accumulates large amounts of homogentisic acid in the urine and has other symptoms of the disease known as alkaptonuria. Similarly, if a person has two mutant (defective) alleles of the gene encoding phenylalanine hydroxylase, he or she is unable to synthesize the enzyme phenylalanine hydroxylase and has the disease called phenylketonuria (PKU).

Concept check: Which disease occurs when homogentisic acid oxidase is defective?

Beadle and Tatum's Experiments with *Neurospora* Led Them to Propose the One-Gene/One-Enzyme Hypothesis

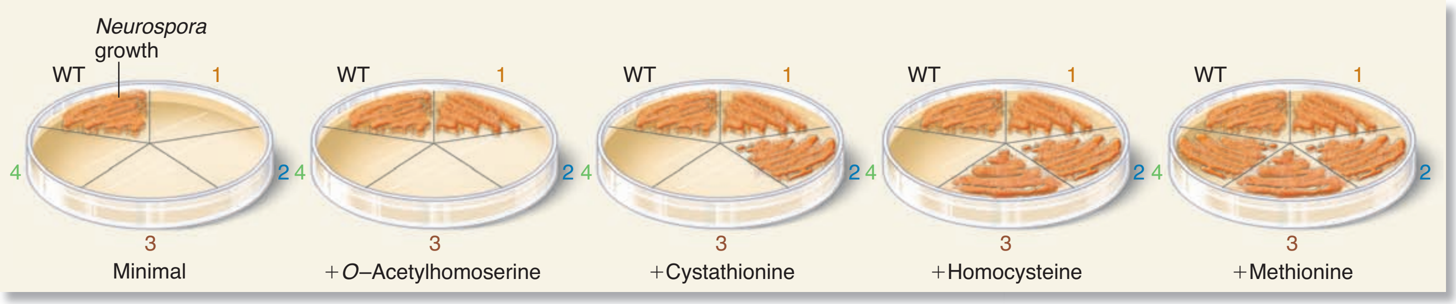
In the early 1940s, George Beadle and Edward Tatum were also interested in the relationship among genes, enzymes, and traits. They developed an experimental system for investigating the connection between genes and the production of particular enzymes. Consistent with the ideas of Garrod, the underlying assumption behind their approach was that a relationship exists between genes and the production of enzymes. However, the quantitative nature of this relationship was unclear. In particular, they asked the question, "Does one gene control the production of one enzyme, or does one gene control the synthesis of many enzymes involved in a complex biochemical pathway?"

At the time of their studies, many geneticists were trying to understand the nature of the gene by studying morphological traits. However, Beadle and Tatum realized that morphological traits are likely to be based on systems of biochemical reactions so complex as to make analysis exceedingly difficult. Therefore, they turned their genetic studies to the analysis of simple nutritional requirements in *Neurospora crassa*, a common bread mold. *Neurospora* can be easily grown in the laboratory and has few nutritional requirements: a carbon source (sugar), inorganic salts, and the vitamin biotin. Normal *Neurospora* cells produce many different enzymes that can synthesize the organic molecules, such as amino acids and other vitamins, which are essential for growth.

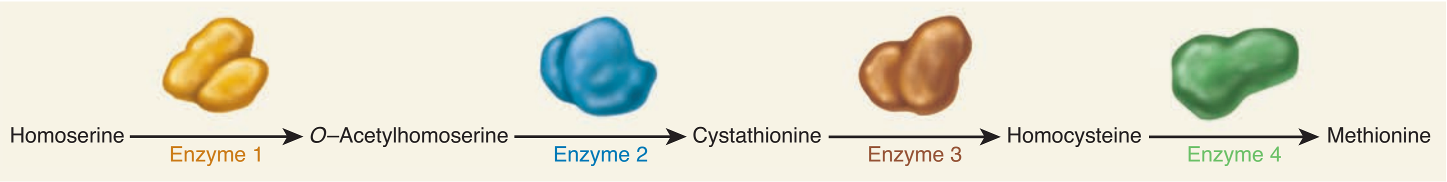
Beadle and Tatum wanted to understand how enzymes are controlled by genes. They reasoned that a mutation in a gene, causing a defect in an enzyme needed for the cellular synthesis of an essential molecule, would prevent that mutant strain from growing on minimal medium, which contains only a carbon source, inorganic salts, and biotin. In their original study of 1941, Beadle and Tatum exposed *Neurospora* cells to X-rays, which caused mutations to occur, and studied the growth of the resulting cells. By plating the cells on media with or without vitamins, they were able to identify mutant strains that required vitamins for growth. In each case, a single mutation resulted in the requirement for a single type of vitamin in the growth media.

This early study by Beadle and Tatum led to additional research to study enzymes involved with the synthesis of other substances, including the amino acid methionine. They first isolated several different mutant strains that required methionine for growth. They hypothesized that each mutant strain might be blocked at only a single step in the consecutive series of reactions that lead to methionine synthesis. To test this hypothesis, the mutant strains were examined for their ability to grow in the presence of *O*-acetylhomoserine, cystathionine, homocysteine, or methionine. *O*-Acetylhomoserine, cystathionine, and homocysteine are intermediates in the synthesis of methionine from homoserine. A simplified depiction of the results is shown in **Figure 15.2a**. The wild-type strain could grow on minimal growth media that contained the minimum set of nutrients that is required for growth. The minimal media did not contain *O*-acetylhomoserine, cystathionine, homocysteine, or methionine. Based on their growth properties on media supplemented with *O*-acetylhomoserine, cystathionine, homocysteine, or methionine, the mutant strains that had been originally identified as requiring methionine for growth could be placed into four groups designated strains 1, 2, 3, and 4 in this figure. A strain 1 mutant was missing enzyme 1, needed for the conversion of homoserine into *O*-acetylhomoserine. The cells could grow only if *O*-acetylhomoserine, cystathionine, homocysteine, or methionine was added to the growth medium. A strain 2 mutant was missing the second enzyme in this pathway that is needed for the conversion of *O*-acetylhomoserine into cystathionine, and a strain 3 mutant was unable to convert cystathionine into homocysteine. Finally, a strain 4 mutant could not make methionine from homocysteine. Based on these results, the researchers could order the enzymes into a biochemical pathway as depicted in **Figure 15.2b**. Taken together, the analysis of these mutants allowed Beadle and Tatum to conclude that a single gene controlled the synthesis of a single enzyme. This was referred to as the **one-gene/one-enzyme hypothesis.**

In later decades, this hypothesis had to be modified in three ways. First, enzymes are only one category of cellular proteins.



(a) Growth of strains on minimal and supplemented growth media



(b) Simplified pathway for methionine biosynthesis

FIGURE 15.2 **An example of an experiment that supported Beadle and Tatum's one-gene/one-enzyme hypothesis.** (**a**) Growth of wild-type (WT) and mutant strains on minimal growth media or in the presence of *O*-acetylhomoserine, cystathionine, homocysteine, or methionine. (**b**) A simplified pathway for methionine biosynthesis. Note: Homoserine is made by *Neurospora* via enzymes and precursor molecules not discussed in this experiment.

Concept check: *What enzymatic function is missing in the strain 2 mutants?*

All proteins are encoded by genes, and many of them do not function as enzymes. Second, some proteins are composed of two or more different polypeptides. The term **polypeptide** refers to a structure; it is a linear sequence of amino acids. By comparison, the term **protein** denotes function. Some proteins are composed of one polypeptide. In such cases, a single gene does encode a single protein. In other cases, however, a functional protein is composed of two or more different polypeptides. An example is hemoglobin, which is composed of two α-globin and two β-globin polypeptides. In this case, two genes—the α-globin and β-globin genes—must be expressed for one functional protein to be made. Therefore, it is more accurate to restate this as the one-gene/one-polypeptide hypothesis. A third reason why the one-gene/one-enzyme hypothesis needed revision is that we now know that many genes do not encode polypeptides. As discussed in Chapter 14, several types of genes specify functional RNA molecules that are not used to encode polypeptides (refer back to Table 14.1).

15.1 REVIEWING THE KEY CONCEPTS

- Garrod studied the disease called alkaptonuria and suggested that some genes encode enzymes (see Figure 15.1).
- Beadle and Tatum studied *Neurospora* mutants that were altered in their nutritional requirements and hypothesized that one gene encodes one enzyme. This hypothesis was later modified because: (1) some proteins are not enzymes; (2) some proteins are composed of two or more different polypeptides; and (3) some genes encode RNAs that are not translated into polypeptides (see Figure 15.2).

15.1 COMPREHENSION QUESTIONS

1. An inborn error of metabolism is caused by
   a. a mutation in a gene that causes an enzyme to be inactive.
   b. a mutation in a gene that occurs in somatic cells.
   c. the consumption of foods that disrupt metabolic processes.
   d. all of the above.
2. The reason why Beadle and Tatum observed four different categories of mutants that could not grow on media without methionine is because
   a. the enzyme involved in methionine biosynthesis is composed of four different subunits.
   b. the enzyme involved in methionine biosynthesis is present in four copies in the *Neurospora* genome.
   c. four different enzymes are involved in a pathway for methionine biosynthesis.
   d. a lack of methionine biosynthesis can inhibit *Neurospora* growth in four different ways.

15.2 THE RELATIONSHIP BETWEEN THE GENETIC CODE AND PROTEIN SYNTHESIS

Thus far, we have considered experiments that led to the conclusion that some genes encode enzymes. Further research indicated that most genes encode polypeptides that form functional units within proteins. The sequence of a structural gene provides a template for the synthesis of mRNA, which in turn contains the information to synthesize a polypeptide. In this section, we will examine the general features of the genetic code—the sequence of bases in a codon that specifies an amino acid. In addition, we will look at the biochemistry of polypeptide synthesis and explore the structure and function of proteins. Ultimately, proteins are largely responsible for determining the characteristics of living cells and an organism's traits.

During Translation, the Genetic Code Within mRNA Is Used to Make a Polypeptide with a Specific Amino Acid Sequence

Why have researchers given the name *translation* to the process of polypeptide synthesis? At the molecular level, translation involves an interpretation of one language—the language of mRNA, a nucleotide sequence—into the language of proteins—an amino acid sequence. The ability of mRNA to be translated into a specific sequence of amino acids relies on the **genetic code.** The sequence of bases within an mRNA molecule provides coded information that is read in groups of three nucleotides known as codons (**Figure 15.3**). The sequence of three bases in most codons specifies a particular amino acid. These codons are termed **sense codons.** For example, the codon AGC specifies the amino acid serine, whereas the codon GGG encodes the amino acid glycine. The codon AUG, which specifies methionine, is used as a **start codon;** it is usually the first codon that begins a polypeptide sequence. The AUG codon can also be used to specify additional methionines within the coding sequence. Finally, three codons —UAA, UAG, and UGA—are used to end the process of translation and are known as **stop codons.** They are also called **termination** or **nonsense codons.** An mRNA molecule also has regions that precede the start codon and follow the stop codon. Because these regions do not encode a polypeptide, they are called the 5′-untranslated region and 3′-untranslated region, respectively.

The codons in mRNA are recognized by the anticodons in **transfer RNA (tRNA)** molecules (see Figure 15.3). **Anticodons** are three-nucleotide sequences that are complementary to codons in mRNA. The tRNA molecules carry the amino acids that correspond to the codons in the mRNA. In this way, the order of codons in mRNA dictates the order of amino acids within a polypeptide.

The details of the genetic code are shown in **Table 15.1.** Because polypeptides are usually composed of 20 different kinds of amino acids, a minimum of 20 codons is needed to specify each type. With four types of bases in mRNA (A, U, G, and C), a genetic code containing two bases in a codon would not be sufficient because it would have only 4^2, or 16, possible types. By comparison, a three-base codon system can specify 4^3, or 64, different codons. Because the number of possible codons exceeds 20—which is the number of different types of amino acids—the genetic code is termed **degenerate.** This means that more than one codon can specify the same amino acid. For example, the codons GGU, GGC, GGA, and GGG all specify the amino acid

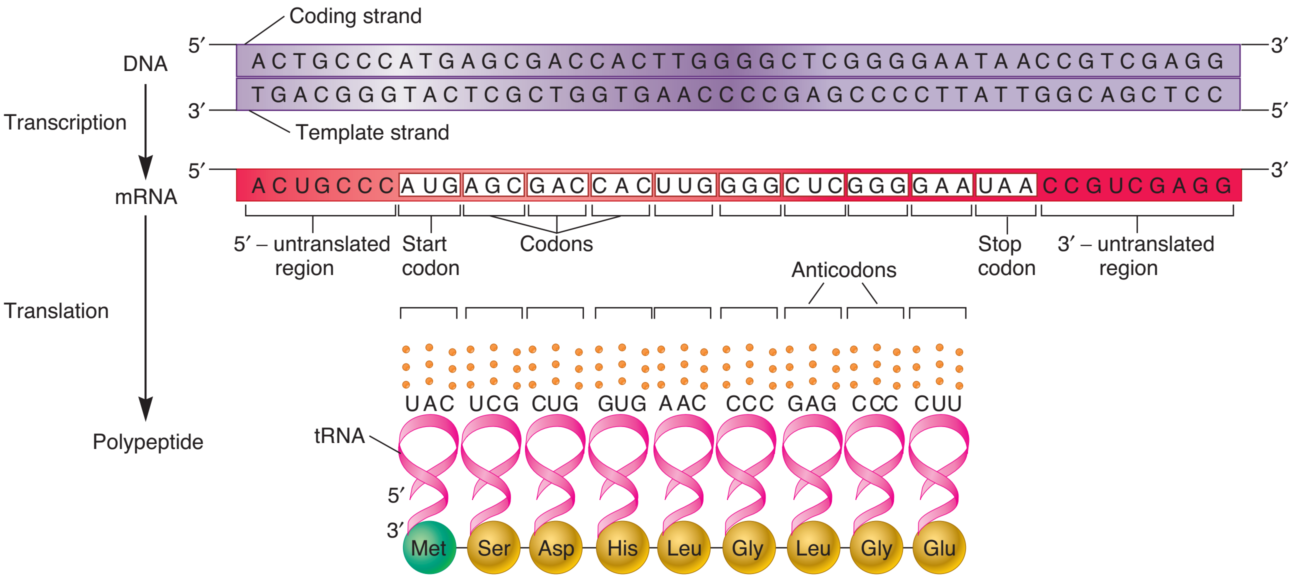

Coding strand
DNA
5′
ACTGCCCATGAGCGACCACTTGGGGCTCGGGGAATAACCGTCGAGG
3′
3′
TGACGGGTACTCGCTGGTGAACCCCGAGCCCCTTATTGGCAGCTCC
5′
Template strand
Transcription
mRNA
5′
ACUGCCC AUG AGC GAC CAC UUG GGG CUC GGG GAA UAA CCGUCGAGG
3′
5′ – untranslated region
Start codon
Codons
Stop codon
3′ – untranslated region
Anticodons
Translation
UAC UCG CUG GUG AAC CCC GAG CCC CUU
tRNA
5′
3′
Polypeptide
Met Ser Asp His Leu Gly Leu Gly Glu

FIGURE 15.3 **The relationships among the DNA-coding sequence, mRNA codons, tRNA anticodons, and amino acids in a polypeptide.** The sequence of nucleotides within DNA is transcribed to make a complementary sequence of nucleotides within mRNA. This sequence of nucleotides in mRNA is translated into a sequence of amino acids of a polypeptide. tRNA molecules act as intermediates in this translation process.

Concept check: *Describe the role of DNA in the synthesis of a polypeptide.*

glycine. Such codons are termed **synonymous codons.** In most instances of synonymous codons, the third base in the codon is the base that varies. The third base is sometimes referred to as the **wobble base.** This term is derived from the idea that the complementary base in the tRNA can "wobble" a bit during the recognition of the third base of the codon in mRNA. The significance of the wobble base will be discussed later in this chapter.

The start codon (AUG) defines the **reading frame** of an mRNA—a sequence of codons determined by reading bases in groups of three, beginning with the start codon. This concept is best understood with a few examples. The following mRNA sequence encodes a short polypeptide with seven amino acids (see next page):

TABLE 15.1

The Genetic Code

Second base

First base	U	C	A	G	Third base
U	UUU Phenylalanine (Phe)	UCU Serine (Ser)	UAU Tyrosine (Tyr)	UGU Cysteine (Cys)	U
	UUC	UCC	UAC	UGC	C
	UUA Leucine (Leu)	UCA	UAA Stop codon	UGA Stop codon	A
	UUG	UCG	UAG Stop codon	UGG Tryptophan (Trp)	G
C	CUU Leucine (Leu)	CCU Proline (Pro)	CAU Histidine (His)	CGU Arginine (Arg)	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Glutamine (Gln)	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Isoleucine (Ile)	ACU Threonine (Thr)	AAU Asparagine (Asn)	AGU Serine (Ser)	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lysine (Lys)	AGA Arginine (Arg)	A
	AUG Methionine (Met); start codon	ACG	AAG	AGG	G
G	GUU Valine (Val)	GCU Alanine (Ala)	GAU Aspartic acid (Asp)	GGU Glycine (Gly)	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glutamic acid (Glu)	GGA	A
	GUG	GCG	GAG	GGG	G

5′–AUGCCCGGAGGCACCGUCCAAU–3′
Met–Pro–Gly–Gly–Thr–Val–Gln

If we remove one base (C) adjacent to the start codon, this changes the reading frame to produce a different polypeptide sequence:

5′–AUGCCGGAGGCACCGUCCAAU–3′
Met–Pro–Glu–Ala–Pro–Ser–Asn

Alternatively, if we remove three bases (CCC) next to the start codon, the resulting polypeptide has the same reading frame as the first polypeptide, though one amino acid (Pro, proline) has been deleted:

5′–AUGGGAGGCACCGUCCAAU–3′
Met–Gly–Gly–Thr–Val–Gln

Exceptions to the Genetic Code Are Known to Occur, Which Includes the Incorporation of Selenocysteine and Pyrrolysine into Polypeptides

From the analysis of many different species, including bacteria, archaea, protists, fungi, plants, and animals, researchers have found that the genetic code is nearly universal. This means that most features of the genetic code are used by all living organisms. However, a few exceptions to the genetic code have been noted (**Table 15.2**). The eukaryotic organelles known as mitochondria have their own DNA, which encodes a few structural genes. In mammals, the mitochondrial genetic code contains differences such as AUA = methionine and UGA = tryptophan. Also, in mitochondria and certain ciliated protists, AGA and AGG specify stop codons instead of arginine.

Selenocysteine (Sec) and **pyrrolysine** (Pyl) are sometimes called the 21st and 22nd amino acids in polypeptides. Their structures are shown later in Figure 15.5f. Selenocysteine is found in several enzymes involved in oxidation-reduction reactions in bacteria, archaea, and eukaryotes. Pyrrolysine is found in a few enzymes of methane-producing archaea. Selenocysteine and pyrrolysine are encoded by UGA and UAG codons, respectively, which normally function as stop codons. Like the standard 20 amino acids, selenocysteine and pyrrolysine are bound to tRNAs that specifically carry them to the ribosome for their incorporation into polypeptides. The anticodon of the tRNA that carries selenocysteine is complementary to a UGA codon, and the tRNA that carries pyrrolysine has an anticodon that is complementary to UAG.

TABLE **15.2**

Examples of Exceptions to the Genetic Code*

Codon	Universal Meaning	Exception
AUA	Isoleucine	Methionine in yeast and mammalian mitochondria
UGA	Stop	Tryptophan in mammalian mitochondria
CUU, CUA, CUC, CUG	Leucine	Threonine in yeast mitochondria
AGA, AGG	Arginine	Stop codon in ciliated protozoa and in yeast and mammalian mitochondria
UAA, UAG	Stop	Glutamine in ciliated protozoa
UGA	Stop	Encodes selenocysteine in certain genes found in bacteria, archaea, and eukaryotes
UAC	Stop	Encodes pyrrolysine in certain genes found in methane-producing archaea

*Several other exceptions, sporadically found among various species, are also known.

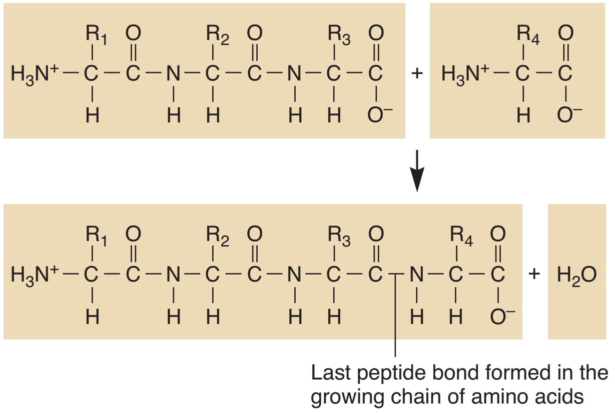
How are UGA and UAG codons "recoded" to specify the incorporation of selenocysteine or pyrrolysine, respectively? In the case of selenocysteine, a UGA codon is followed by a sequence called the selenocysteine insertion sequence (SECIS), which forms a stem-loop. In bacteria, a SECIS may be located immediately following the UGA codon, whereas a SECIS may be further downstream in the 3′-untranslated region of the mRNA in archaea and eukaryotes. The SECIS is recognized by proteins that favor the binding of a UGA codon to a tRNA carrying selenocysteine instead of the binding of release factors that are needed for polypeptide termination. Similarly, pyrrolysine incorporation may involve sequences downstream from a UAG codon that form a stem-loop.

A Polypeptide Has Directionality from Its Amino-Terminal to Its Carboxyl-Terminal End

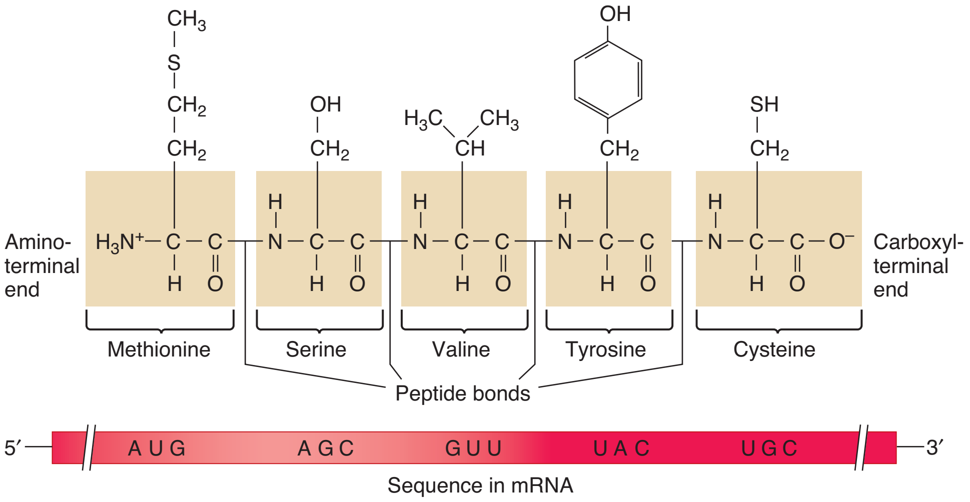
Let's now turn our attention to polypeptide biochemistry. Polypeptide synthesis has a directionality that parallels the order of codons in the mRNA. As a polypeptide is made, a **peptide bond** is formed between the carboxyl group in the last amino acid of the polypeptide and the amino group in the amino acid being added. As shown in **Figure 15.4a**, this occurs via a condensation reaction that releases a water molecule. The newest amino acid added to a growing polypeptide always has a free carboxyl group. **Figure 15.4b** compares the sequence of a very short polypeptide with the mRNA that encodes it. The first amino acid is said to be at the **amino-terminal end,** or **N-terminus,** of the polypeptide. An amino group (NH_3^+) is found at this site. The term *N-terminus* refers to the presence of a nitrogen atom (N) at this end. The first amino acid is specified by a codon that is near the 5′ end of the mRNA. By comparison, the last amino acid in a completed polypeptide is located at the **carboxyl-terminal end,** or **C-terminus.** A carboxyl group (COO^-) is always found at this site in a polypeptide. This last amino acid is specified by a codon that is closer to the 3′ end of the mRNA.

The Amino Acid Sequences of Polypeptides Determine the Structure and Function of Proteins

Now that we understand how mRNAs encode polypeptides, let's consider the structure and function of the gene product, namely, polypeptides. **Figure 15.5** shows the 22 different amino acids that may be found within polypeptides. Each amino acid contains a unique **side chain,** or **R group,** that has its own particular chemical properties. For example, aliphatic and aromatic amino acids are relatively nonpolar, which means they are less likely to


R1 O
R2 O
R3 O
R4 O
H3N+
+
H2O
Last peptide bond formed in the growing chain of amino acids

(a) Attachment of an amino acid to a peptide chain



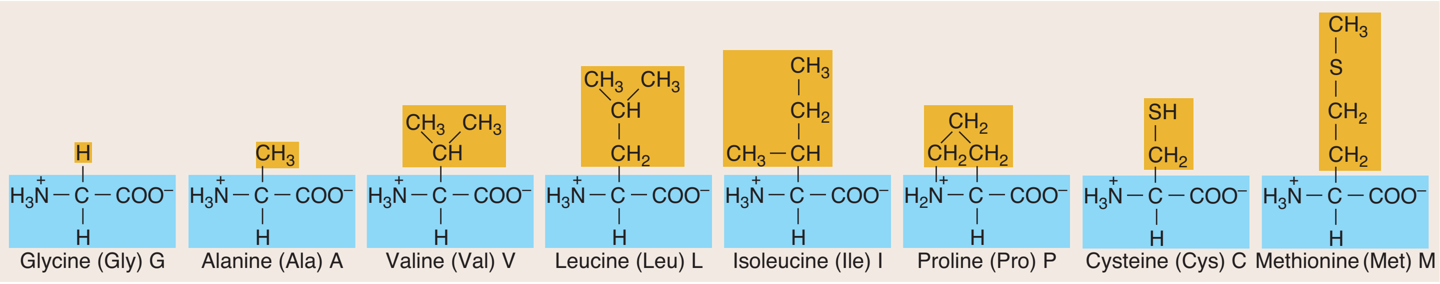
(b) Directionality in a polypeptide and mRNA

FIGURE 15.4 **The directionality of polypeptide synthesis.** **(a)** An amino acid is connected to a polypeptide via a condensation reaction that releases a water molecule. The letter R is a general designation for an amino acid side chain. **(b)** The first amino acid in a polypeptide (usually methionine) is located at the amino-terminal end, and the last amino acid is at the carboxyl-terminal end. Therefore, the directionality of amino acids in a polypeptide is from the amino-terminal end to the carboxyl-terminal end, which corresponds to the 5′ to 3′ orientation of codons in mRNA.

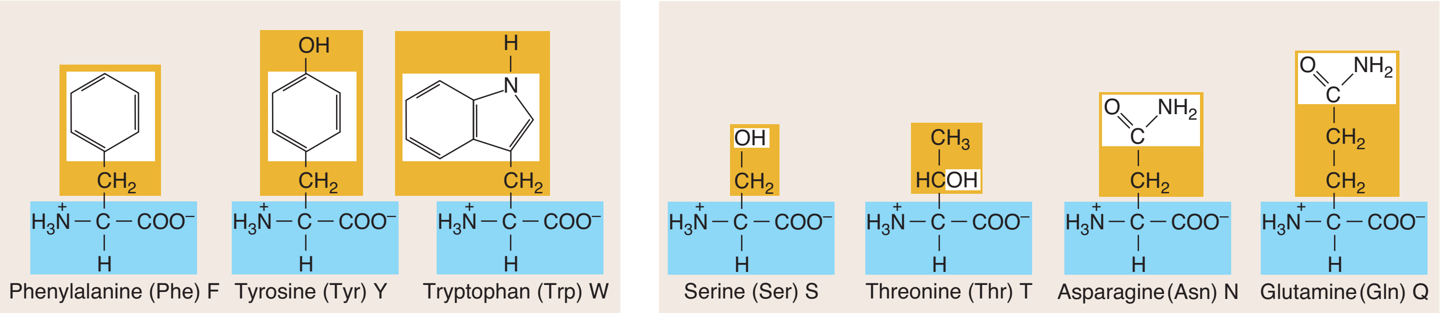
associate with water. These hydrophobic (meaning water-fearing) amino acids are often buried within the interior of a folded protein. In contrast, the polar amino acids are hydrophilic (water-loving) and are more likely to be on the surface of a protein, where they can favorably interact with the surrounding water. The chemical properties of the amino acids and their sequences in a polypeptide are critical factors that determine the unique structure of that polypeptide.

Following gene transcription and mRNA translation, the end result is a polypeptide with a defined amino acid sequence. This sequence is the **primary structure** of a polypeptide. Figure 15.6 shows the primary structure of an enzyme called lysozyme, a relatively small protein containing 129 amino acids. The primary structure of a typical polypeptide may be a few hundred or even a couple of thousand amino acids in length. Within a living cell, a newly made polypeptide is not usually found in a long linear state for a significant length of time. Rather, to become a functional unit, most polypeptides quickly adopt a compact three-dimensional structure. The folding process begins while the polypeptide is still being translated. The progression from the primary structure of a polypeptide to the three-dimensional structure of a protein is dictated by the amino acid sequence within the polypeptide. In particular, the chemical properties of the amino acid side chains play a central role in determining the folding pattern of a protein. In addition, the folding of some polypeptides is aided by **chaperones**—proteins that bind to polypeptides and facilitate their proper folding.

This folding process of polypeptides is governed by the primary structure and occurs in multiple stages (Figure 15.7). The first stage involves the formation of a regular, repeating shape known as a **secondary structure.** The two main types of secondary structures are the **α helix** and the **β sheet** (Figure 15.7b). A single polypeptide may have some regions that fold into an α helix and other regions that fold into a β sheet. Because of the

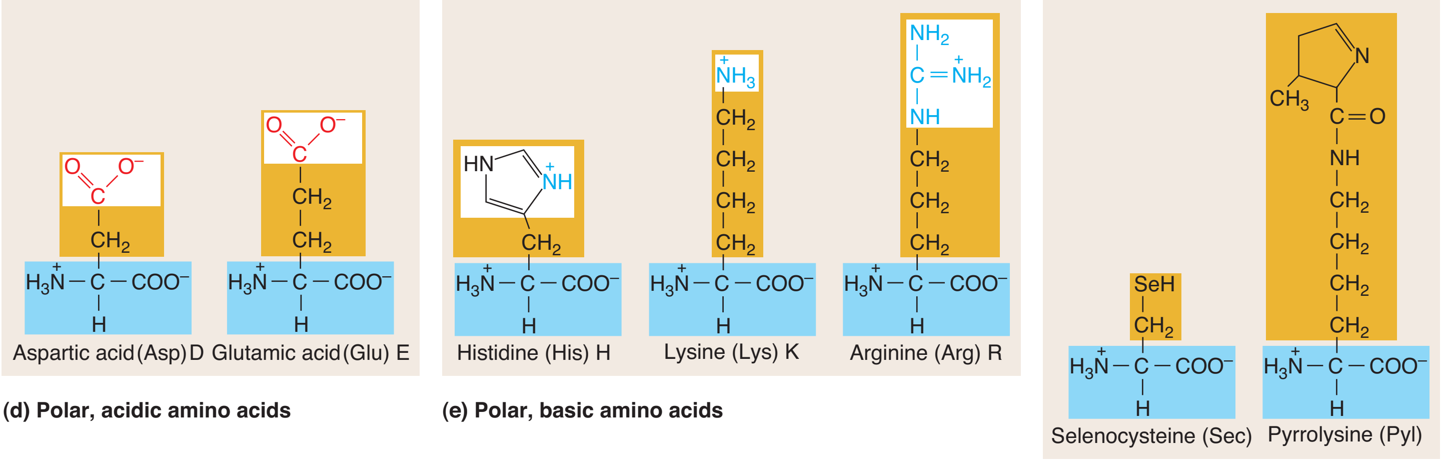

Glycine (Gly) G
Alanine (Ala) A
Valine (Val) V
Leucine (Leu) L
Isoleucine (Ile) I
Proline (Pro) P
Cysteine (Cys) C
Methionine (Met) M

(a) Nonpolar, aliphatic amino acids


Phenylalanine (Phe) F
Tyrosine (Tyr) Y
Tryptophan (Trp) W
Serine (Ser) S
Threonine (Thr) T
Asparagine (Asn) N
Glutamine (Gln) Q

(b) Aromatic amino acids

(c) Polar, neutral amino acids


Aspartic acid (Asp) D
Glutamic acid (Glu) E
Histidine (His) H
Lysine (Lys) K
Arginine (Arg) R
Selenocysteine (Sec)
Pyrrolysine (Pyl)

(d) Polar, acidic amino acids

(e) Polar, basic amino acids

(f) Nonstandard amino acids

FIGURE 15.5 **The amino acids that are incorporated into polypeptides during translation.** Parts (a) through (e) show the 20 standard amino acids, and part (f) shows two amino acids that are occasionally incorporated into polypeptides by the use of stop codons (see Table 15.2). The structures of amino acid side chains can also be covalently modified after a polypeptide is made, a phenomenon called post-translational modification.

Concept check: *Which two amino acids do you think are the least soluble in water?*

geometry of secondary structures, certain amino acids, such as glutamic acid, alanine, and methionine, are good candidates to form an α helix. Other amino acids, such as valine, isoleucine, and tyrosine, are more likely to be found in a β-sheet conformation. Secondary structures within polypeptides are primarily stabilized by the formation of hydrogen bonds between atoms that are located in the polypeptide backbone. In addition, some regions do not form a repeating secondary structure. Such regions have shapes that look very irregular in their structure because they do not follow a repeating folding pattern.

The short regions of secondary structure within a polypeptide are folded relative to each other to make the **tertiary structure** of a polypeptide. As shown in Figure 15.7c, α-helical regions and β-sheet regions are connected by irregularly shaped segments to determine the tertiary structure of the polypeptide. The folding of a polypeptide into its secondary and then tertiary conformation can usually occur spontaneously because it is a thermodynamically favorable process. The structure is determined by various interactions, including the tendency of hydrophobic amino acids to avoid water, ionic interactions among charged amino acids, hydrogen bonding among amino acids in the folded polypeptide, and weak bonding known as van der Waals interactions.

A protein is a functional unit that can be composed of one or more polypeptides. Some proteins are composed of a single polypeptide. Many proteins, however, are composed of two or more polypeptides that associate with each other to make a functional protein with a **quaternary structure** (Figure 15.7d). The individual polypeptides are called **subunits** of the protein, each of which has its own tertiary structure.

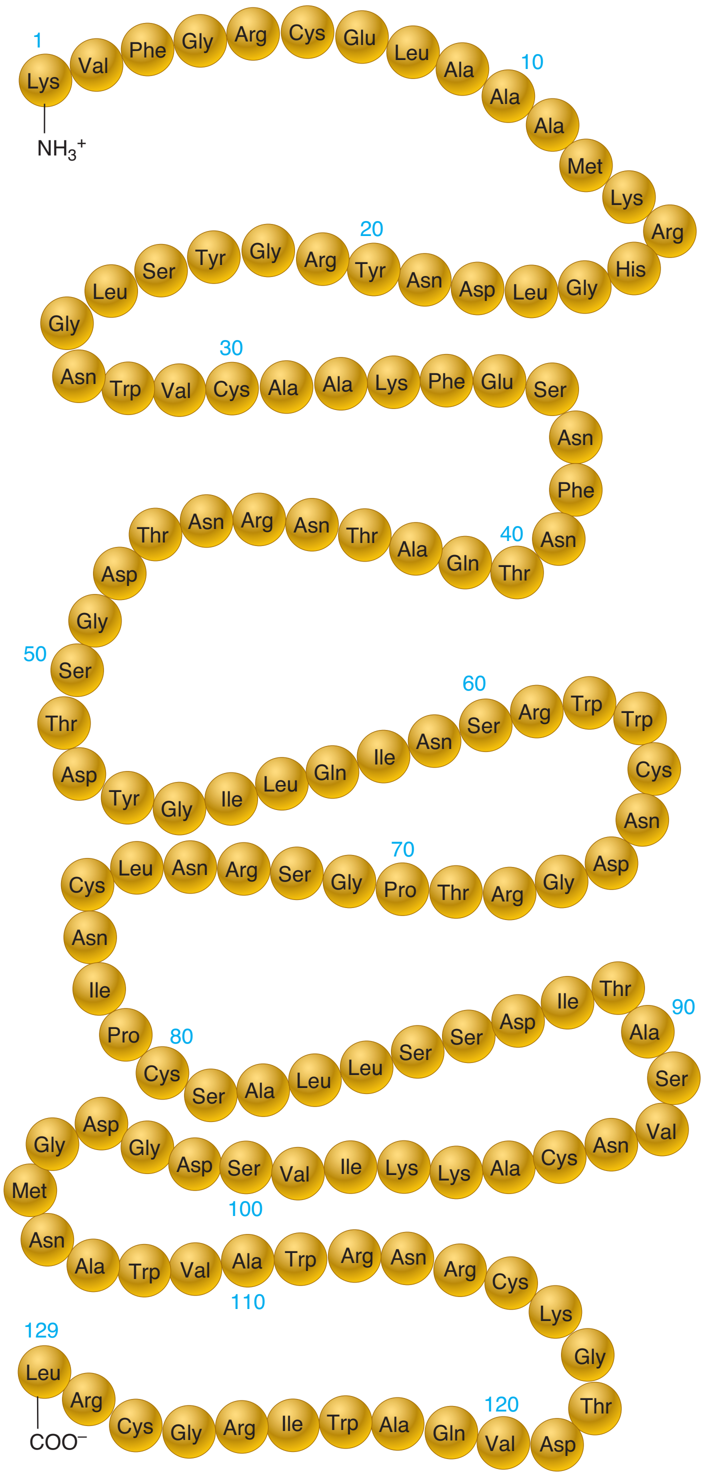

1
Lys
NH$_3^+$
Val
Phe
Gly
Arg
Cys
Glu
Leu
Ala
10
Ala
Ala
Met
Lys
Arg
His
Gly
Leu
Asp
Asn
20
Tyr
Arg
Gly
Tyr
Ser
Leu
Gly
Asn
Trp
Val
30
Cys
Ala
Ala
Lys
Phe
Glu
Ser
Asn
Phe
Asn
40
Thr
Gln
Ala
Thr
Asn
Arg
Asn
Thr
Asp
Gly
50
Ser
Thr
Asp
Tyr
Gly
Ile
Leu
Gln
Ile
Asn
60
Ser
Arg
Trp
Trp
Cys
Asn
Asp
Gly
Arg
Thr
70
Pro
Gly
Ser
Arg
Asn
Leu
Cys
Asn
Ile
Pro
80
Cys
Ser
Ala
Leu
Leu
Ser
Ser
Asp
Ile
Thr
90
Ala
Ser
Val
Asn
Cys
Ala
Lys
Lys
Ile
Val
100
Ser
Asp
Gly
Asp
Gly
Met
Asn
Ala
Trp
Val
110
Ala
Trp
Arg
Asn
Arg
Cys
Lys
Gly
Thr
Asp
120
Val
Gln
Ala
Trp
Ile
Arg
Gly
Cys
Arg
129
Leu
COO$^-$

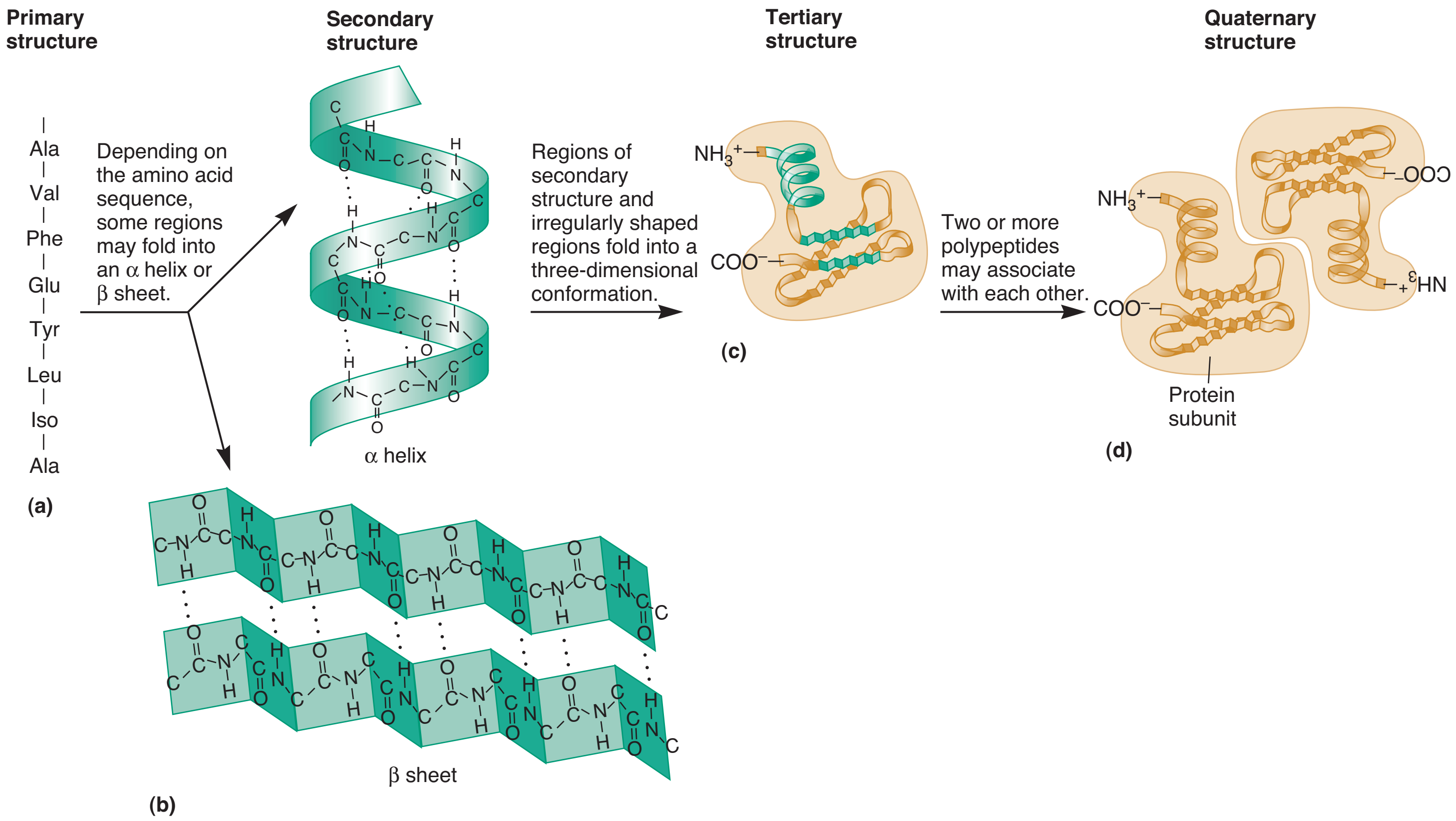
FIGURE 15.6 **An example of a protein's primary structure.** This is the amino acid sequence of the enzyme lysozyme, which contains 129 amino acids in its primary structure. As you may have noticed, the first amino acid is not methionine; instead, it is lysine. The first methionine in this polypeptide sequence is removed after or during translation. The removal of the first methionine occurs in many (but not all) proteins.

15.2 REVIEWING THE KEY CONCEPTS

- During translation, the sequence of codons in mRNA is used via tRNA molecules to make a polypeptide with a specific amino acid sequence (see Figure 15.3).
- The genetic code refers to the relationship between three-base codons in the mRNA and the amino acids that are incorporated into a polypeptide. One codon (AUG) is a start codon, which determines the reading frame of the mRNA. Three codons (UAA, UAG, and UGA) function as stop codons (see Table 15.1).
- The genetic code is largely universal but some exceptions are known to occur (see Table 15.2).
- A polypeptide is made by the formation of peptide bonds between adjacent amino acids. Each polypeptide has a directionality from its amino-terminal end to its carboxyl-terminal end, which parallels the arrangement of codons in mRNA in the 5′ to 3′ direction (see Figure 15.4).
- Amino acids differ in the structure of their side chains (see Figure 15.5).
- Protein structure can be viewed at different levels, which consist of primary structure (sequence of amino acids), secondary structure (repeating folding patterns such as the α helix and the β sheet), tertiary structure (additional folding), and quaternary structure (the binding of multiple subunits to each other) (see Figures 15.6, 15.7).

15.2 COMPREHENSION QUESTIONS

1. What is the genetic code?
    a. The relationship between a three-base codon sequence and an amino acid
    b. The entire base sequence of an mRNA molecule
    c. The entire sequence from the promoter to the terminator of a gene
    d. The binding of tRNA to mRNA
2. The reading frame begins with a _____________ and is read _____________.
    a. promoter, one base at a time
    b. promoter, in groups of three bases
    c. start codon, one base at a time
    d. start codon, in groups of three bases
3. The fourth codon in an mRNA-coding sequence is GGG, which specifies glycine. If we assume that no amino acids are removed from the polypeptide, which of the following statements is *correct*?
    a. The third amino acid from N-terminus is glycine.
    b. The fourth amino acid from the N-terminus is glycine.



Depending on the amino acid sequence, some regions may fold into an α helix or β sheet.

Regions of secondary structure and irregularly shaped regions fold into a three-dimensional conformation.

Two or more polypeptides may associate with each other.

ONLINE ANIMATION

FIGURE 15.7 **Levels of structures formed in proteins.** **(a)** The primary structure of a polypeptide is its amino acid sequence. **(b)** Certain regions of a primary structure fold into a secondary structure; the two types of secondary structures are called α helices and β sheets. **(c)** Both of these secondary structures can be found within the tertiary structure of a polypeptide. **(d)** Some polypeptides associate with each other to form a protein with a quaternary structure.

Concept check: *What type of bonding is responsible for the formation of the two types of secondary structures?*

  c. The third amino acid from the C-terminus is glycine.
  d. The fourth amino acid from the C-terminus is glycine.

4. A type of secondary structure found in proteins is
  a. an α helix.
  b. a β sheet.
  c. both a and b.
  d. none of the above.

15.3 EXPERIMENTAL DETERMINATION OF THE GENETIC CODE

In the previous section, we examined how the genetic code determines the amino acid sequence of a polypeptide. In this section, we will consider the experimental approaches that deduced the genetic code.

Studies Involving T4 Phage Indicated That the Genetic Code Is Read in Triplets

How did researchers discover that the genetic code is read in triplets? The first evidence came from studies of Francis Crick and his colleagues in 1961. These experiments involved the isolation of mutants in a bacteriophage called T4. As described in Chapter 10, mutations in T4 genes that affect plaque morphology are easily identified (see Figure 10.8). In particular, loss-of-function mutations within certain T4 genes, designated *rII,* resulted in plaques that were larger and had a clear boundary. In comparison, wild-type phages, designated r^+, produced smaller plaques with a fuzzy boundary. Crick and colleagues exposed T4 phages to a chemical called proflavin, which causes single-nucleotide additions or deletions in gene sequences. The mutagenized phages were plated to identify large (*rII*) plaques. Though proflavin can cause either single-nucleotide additions or deletions, the first mutant strain that the researchers identified was arbitrarily called a (+) mutation. Many years later, when methods of DNA sequencing became available, it was determined that the (+) mutation is a single-nucleotide addition. Table 15.3 shows a hypothetical wild-type sequence in a phage gene (first line) and considers how nucleotide additions and/or deletions could affect the resulting amino acid sequence. A single-nucleotide addition (+) would alter the reading frame beyond the point of insertion, thereby abolishing the proper function of the encoded protein. This is called a **frameshift mutation,** because it has changed the reading frame. This mutation results in a loss of function for the protein encoded by this viral gene, thereby producing an *rII* plaque phenotype.

TABLE **15.3**

Evidence That the Genetic Code Is Read in Triplets*

Strain	Plaque Phenotype	DNA Coding Sequence/Polypeptide Sequence§	Downstream Sequence‡
Wild type	r^+	ATG GGG CCC GTC CAT CCG TAC GCC GGA ATT ATA Met Gly Pro Val His Pro Tyr Ala Gly Ile Ile---------	In frame
(+)	*rII*	↓A ATG GGG ACC CGT CCA TCC GTA CGC CGG AAT TAT A Met Gly Thr Arg Pro Ser Val Arg Arg Asn Tyr--------	Out of frame
(+)(–)$_a$	r^+	↓A ↑C ATG GGG ACC GTC CAT CCG TAC GCC GGA ATT ATA Met Gly Thr Val His Pro Tyr Ala Gly Ile Ile---------	In frame
(+)(–)$_b$	r^+	↓A ↑T ATG GGG ACC CGC CAT CCG TAC GCC GGA ATT ATA Met Gly Thr Arg His Pro Tyr Ala Gly Ile Ile---------	In frame
(+)(–)$_c$	r^+	↓A ↑G ATG GGG ACC CTC CAT CCG TAC GCC GGA ATT ATA Met Gly Thr Leu His Pro Tyr Ala Gly Ile Ile---------	In frame
(–)$_a$	*rII*	↑C ATG GGG CCG TCC ATC CGT ACG CCG GAA TTA TA Met Gly Pro Ser Ile Arg Thr Pro Glu Leu---------------	Out of frame
(–)$_a$(–)$_b$	*rII*	↑C ↑T ATG GGG CCG CCA TCC GTA CGC CGG AAT TAT A Met Gly Pro Pro Ser Val Arg Arg Asn Tyr----------------	Out of frame
(–)$_a$(–)$_b$(–)$_c$	r^+	↑↑↑GTC ATG GGG CCC CAT CCG TAC GCC GGA ATT ATA Met Gly Pro His Pro Tyr Ala Gly Ile Ile-------------	In frame [Only Val is missing]

*This table shows only a small portion of a hypothetical coding sequence.

§A down arrow (↓) indicates the location of a single nucleotide addition; an up arrow (↑) indicates the location of a single nucleotide deletion.

‡The term *downstream sequence* refers to the remaining part of the sequence that is not shown in this figure. It includes hundreds of codons. An *in-frame sequence* is wild type, whereas an *out-of-frame sequence* (caused by the addition or deletion of one or two base pairs) is not.

The (+) mutant strain was then subjected to a second round of mutagenesis via proflavin. Several plaques were identified that had reverted to a wild-type (r^+) phenotype. By analyzing these strains using methods described in Chapter 10, it was determined that each one contained a second mutation that was close to the original (+) mutation. These second mutations were designated (–) mutations. Three different (–) mutations, designated a, b, and c, were identified. Each of these (–) mutations was a single-nucleotide deletion that was close to the original (+) mutation. Therefore, it restored the reading frame and produced a protein with a nearly normal amino acid sequence.

The critical experiment that suggested the genetic code is read in triplets came by combining different (–) mutations together. Mutations in different phages can be brought together into the same phage via crossing over, as described in Chapter 10. Using such an approach, the researchers constructed strains containing one, two, or three (–) mutations. The results showed that a wild-type plaque morphology was obtained only when three (–) mutations were combined in the same phage (see Table 15.3). The three (–) mutations in the same phage restored the normal reading frame. These results were consistent with the hypothesis that the genetic code is read in multiples of three nucleotides.

Synthetic RNA Helped to Determine the Genetic Code

Having learned that the genetic code is read in triplets, how did scientists determine the functions of the 64 codons of the genetic code? During the early 1960s, three research groups headed by Marshall Nirenberg, Severo Ochoa, and H. Gobind Khorana set out to decipher the genetic code. Though they used different methods, all of these groups used synthetic mRNA in their experimental approaches to "crack the code." We first consider the work of Nirenberg and his colleagues. Prior to their studies, several laboratories had already determined that extracts from bacterial cells, containing a mixture of components including ribosomes, tRNAs, and other factors required for translation, are able to synthesize polypeptides if mRNA and amino acids are added. This mixture is termed a **cell-free translation system** (also called an in vitro translation system). If radiolabeled amino acids are added to a cell-free translation system, the synthesized polypeptides are radiolabeled and easy to detect.

To decipher the genetic code, Nirenberg and colleagues needed to gather information regarding the relationship between mRNA composition and polypeptide composition. To accomplish this goal, they made mRNA molecules of a known base composition, added them to a cell-free translation system, and then analyzed the amino acid composition of the resultant polypeptides. For example, if an mRNA molecule consisted of a string of adenine-containing nucleotides (e.g., 5′–AAAAAAAAAAAAAAAA–3′), researchers could add this polyA mRNA to a cell-free translation system and ask the question, "Which amino acid is specified by a codon that contains only adenine nucleotides?" (As Table 15.1 shows, it is lysine.)

Before discussing the details of this type of experiment, let's consider how the synthetic mRNA molecules were made. To synthesize mRNA, an enzyme known as polynucleotide phosphorylase was used. In the presence of excess ribonucleoside diphosphates, also called nucleoside diphosphates (NDPs), this enzyme catalyzes the covalent linkage of nucleotides to make a polymer of RNA. Because it does not use a template, the order of the nucleotides is random. For example, if only uracil-containing diphosphates (UDPs) are added, then a polyU mRNA (5′–UUUUUUUUUUUUUUUU–3′) is made. If nucleotides containing two different bases, such as uracil and guanine, are added, then the phosphorylase makes a random polymer containing both nucleotides (5´–GGGUGUGUGGUGGGUG–3′). An experimenter can control the amounts of the nucleotides that are added. For example, if 70% G and 30% U are mixed together with polynucleotide phosphorylase, the predicted amounts of the codons within the random polymer are as follows:

Codon Possibilities	*Percentage in the Random Polymer*
GGG	0.7 × 0.7 × 0.7 = 0.34 = 34%
GGU	0.7 × 0.7 × 0.3 = 0.15 = 15%
GUU	0.7 × 0.3 × 0.3 = 0.06 = 6%
UUU	0.3 × 0.3 × 0.3 = 0.03 = 3%
UUG	0.3 × 0.3 × 0.7 = 0.06 = 6%
UGG	0.3 × 0.7 × 0.7 = 0.15 = 15%
UGU	0.3 × 0.7 × 0.3 = 0.06 = 6%
GUG	0.7 × 0.3 × 0.7 = 0.15 = 15%
	100%

The first experiment that demonstrated the ability to synthesize polypeptides from synthetic mRNA was performed by Marshall Nirenberg and J. Heinrich Matthaei in 1961. As shown in **Figure 15.8**, a cell-free translation system was added to 20 different tubes. An mRNA template made via polynucleotide phosphorylase was then added to each tube. In this example, the mRNA was made using 70% G and 30% U. Next, the 20 amino acids were added to each tube, but each tube differed with regard to the type of radiolabeled amino acid. For example, radiolabeled glycine is found in only 1 of the 20 tubes. The tubes were incubated a sufficient length of time to allow translation to occur. The newly made polypeptides were then precipitated onto a filter by treatment with trichloroacetic acid. This step precipitates polypeptides but not individual amino acids. A washing step caused amino acids that had not been incorporated into polypeptides to pass through the filter. Finally, the amount of radioactivity trapped on the filter was determined by liquid scintillation counting.

▶ THE GOAL

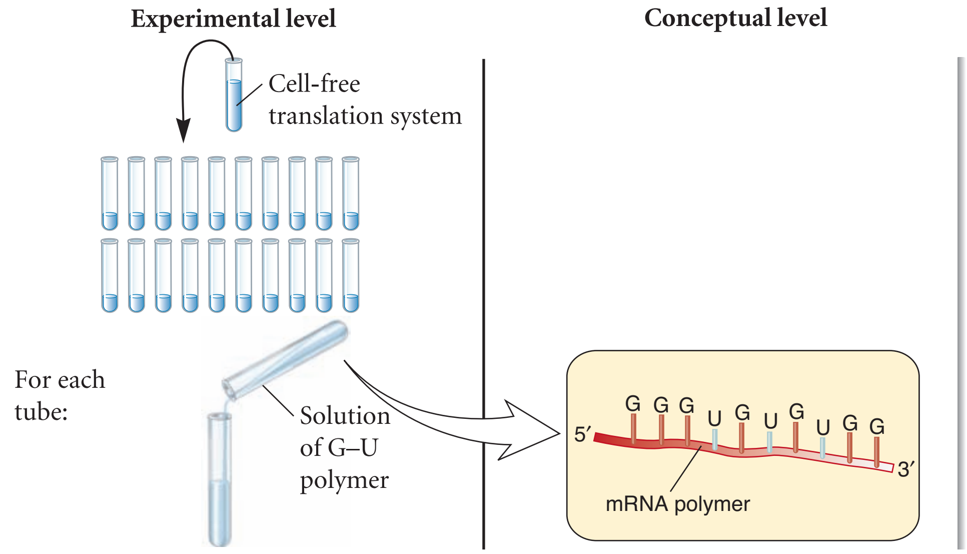
The researchers assumed that the sequence of bases in mRNA determines the incorporation of specific amino acids into a polypeptide. The purpose of this experiment was to provide information to help decipher the relationship between base composition and particular amino acids.

▶ ACHIEVING THE GOAL — FIGURE 15.8 **Elucidation of the genetic code.**

Starting material: A cell-free translation system that can synthesize polypeptides if mRNA and amino acids are added.

1. Add the cell-free translation system to each of 20 tubes.

2. To each tube, add random mRNA polymers of G and U made via polynucleotide phosphorylase using 70% G and 30% U.



3. Add a different radiolabeled amino acid to each tube, and add the other 19 non-radiolabeled amino acids. The translation system contained enzymes (discussed later) that attach amino acids to the appropriate tRNAs.

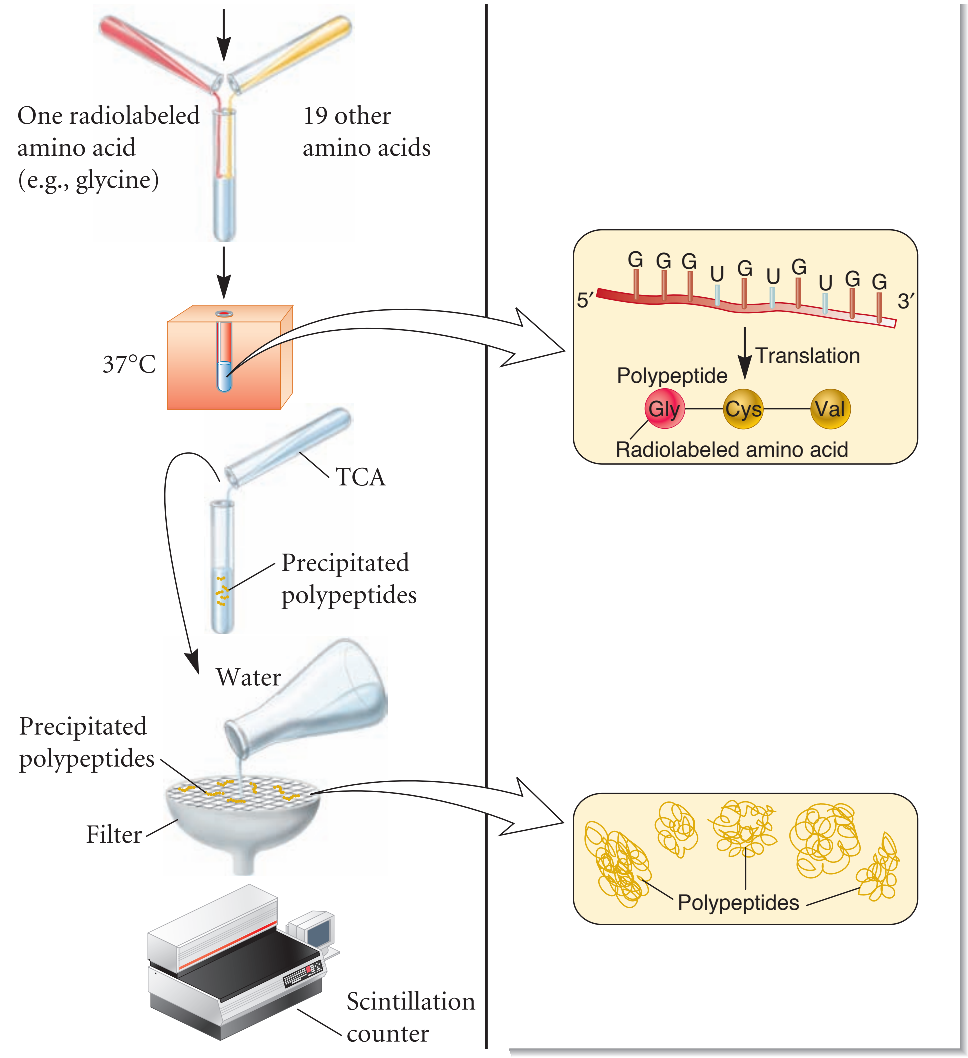
4. Incubate for 60 minutes to allow translation to occur.

5. Add 15% trichloroacetic acid (TCA), which precipitates polypeptides but not amino acids.

6. Place the precipitate onto a filter and wash to remove unused amino acids.

7. Count the radioactivity on the filter in a scintillation counter (see the Appendix for a description).

8. Calculate the amount of radiolabeled amino acids in the precipitated polypeptides.


One radiolabeled amino acid (e.g., glycine)
19 other amino acids
37°C
G G G U G U G U G G
5′
3′
Translation
Polypeptide
Gly
Cys
Val
Radiolabeled amino acid
TCA
Precipitated polypeptides
Water
Precipitated polypeptides
Filter
Polypeptides
Scintillation counter

▶ THE DATA

Radiolabeled Amino Acid Added	*Relative Amount of Radiolabeled Amino Acid Incorporated into Translated Polypeptides (% of total)*
Alanine	0
Arginine	0
Asparagine	0
Aspartic acid	0
Cysteine	6
Glutamic acid	0
Glutamine	0
Glycine	49
Histidine	0
Isoleucine	0
Leucine	6
Lysine	0
Methionine	0
Phenylalanine	3
Proline	0
Serine	0
Threonine	0
Tryptophan	15
Tyrosine	0
Valine	21

Adapted from Marshall W. Nirenberg, and J.H. Matthaei (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA 47*, 1588–1602.

▶ INTERPRETING THE DATA

The genetic code was not deciphered in a single experiment such as the one described here. Furthermore, this kind of experiment yields information regarding only the nucleotide content of codons, not the specific order of bases within a single codon. According to the calculation previously described, codons should occur in the following percentages: 34% GGG, 15% GGU, 6% GUU, 3% UUU, 6% UUG, 15% UGG, 6% UGU, and 15% GUG. Because 6% of the radiolabeling was due to the incorporation of cysteine, these results are consistent with the idea that a cysteine codon contains two U's and one G. However, this single experiment does not tell us that a cysteine codon is UGU. Based on these data alone, a cysteine codon could be UUG, GUU, or UGU. By comparing many different RNA polymers, the laboratories of Nirenberg and Ochoa established patterns between the specific base sequences of codons and the amino acids they encode. In their first experiments, Nirenberg and Matthaei showed that a random polymer containing only uracil produced a polypeptide containing only phenylalanine. From this result, they inferred that UUU specifies phenylalanine. This idea is consistent with the results shown in the data table. In the random 70% G and 30% U polymer, 3% of the codons are UUU. Likewise, 3% of the amino acids within the polypeptides were found to be phenylalanine. Also, in some cases, the percentage of a given amino acid is due to multiple codons. We now know that the value of 49% for glycine is due to two codons: GGG (34%) and GGU (15%).

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

The Use of RNA Copolymers and the Triplet-Binding Assay Also Helped to Crack the Genetic Code

In the 1960s, H. Gobind Khorana and colleagues developed a novel method to synthesize RNA. They first created short RNA molecules, two to four nucleotides in length, that had a defined sequence. For example, RNA molecules with the sequence 5′–UG–3′ were synthesized chemically. These short RNAs were then linked together enzymatically, in a 5′ to 3′ manner, to create long copolymers with the sequence

5′–UGUGUGUGUGUGUGUGUGUGUGUGUGUG–3′

This is called a copolymer, because it is made from the linkage of several smaller molecules. Depending on the reading frame, such a copolymer contains two different codons: UGU and GUG. Using a cell-free translation system like the one described in Figure 15.8, this copolymer produced polypeptides containing cysteine and valine. If these results were compared with the experimental approach of Nirenberg, which showed that a cysteine codon is composed of two U's and one G (see Figure 15.8), the results indicate that UGU specifies cysteine and GUG specifies valine. **Table 15.4** summarizes some of the copolymers that were made using this approach and the amino acids that were incorporated into polypeptides.

Finally, another method that helped to decipher the genetic code also involved the chemical synthesis of short RNA molecules. In 1964, Marshall Nirenberg and Philip Leder discovered that RNA molecules containing three nucleotides—a triplet—could stimulate ribosomes to bind a tRNA. In other words, the RNA triplet acted like a codon. Ribosomes were able to bind RNA triplets, and then a tRNA with the appropriate anticodon could subsequently bind to the ribosome. To establish the relationship between triplet sequences and specific amino acids, samples containing ribosomes and a particular triplet were exposed to tRNAs with different radiolabeled amino acids.

TABLE **15.4**

Examples of Copolymers That Were Analyzed by Khorana and Colleagues

Synthetic RNA*	Codon Possibilities	Amino Acids Incorporated into Polypeptides
UC	UCU, CUC	Serine, leucine
AG	AGA, GAG	Arginine, glutamic acid
UG	UGU, GUG	Cysteine, valine
AC	ACA, CAC	Threonine, histidine
UUC	UUC, UCU, CUU	Phenylalanine, serine, leucine
AAG	AAG, AGA, GAA	Lysine, arginine, glutamic acid
UUG	UUG, UGU, GUU	Leucine, cysteine, valine
CAA	CAA, AAC, ACA	Glutamine, asparagine, threonine
UAUC	UAU, AUC, UCU, CUA	Tyrosine, isoleucine, serine, leucine
UUAC	UUA, UAC, ACU, CUU	Leucine, tyrosine, threonine

*The synthetic RNAs were linked together to make copolymers.

As an example, in one experiment the researchers began with a sample of ribosomes that were mixed with 5′–CCC–3′ triplets. Portions of this sample were then added to 20 different tubes that had tRNAs with different radiolabeled amino acids. For example, one tube contained radiolabeled histidine, a second tube had radiolabeled proline, a third tube contained radiolabeled glycine, and so on. Only one radiolabeled amino acid was added to each tube. After allowing sufficient time for tRNAs to bind to the ribosomes, the samples were filtered; only the large ribosomes and anything bound to them were trapped on the filter (**Figure 15.9**). Unbound tRNAs passed through the filter. Next, the researchers determined the amount of radioactivity trapped on each filter. If the filter contained a large amount of radioactivity, the results indicated that the added triplet encoded the amino acid that was radiolabeled.

Using the triplet-binding assay, Nirenberg and Leder were able to establish relationships between particular triplet sequences and the binding of tRNAs carrying specific (radiolabeled) amino acids. In the case of the 5′–CCC–3′ triplet, they determined that tRNAs carrying radiolabeled proline were bound to the ribosomes. Unfortunately, in some cases, a triplet could not promote sufficient tRNA binding to yield unambiguous results. Nevertheless, the triplet-binding assay was an important tool in the identification of the majority of codons.

15.3 REVIEWING THE KEY CONCEPTS

- Crick studied mutations in T4 phage and determined that the genetic code is read in triplets (see Table 15.3).
- Nirenberg and colleagues used synthetic RNA and a cell-free translation system to decipher the genetic code (see Figure 15.8).
- Other methods of deciphering the genetic code included the synthesis of copolymers by Khorana and the triplet-binding assay of Nirenberg and Leder (see Table 15.4, Figure 15.9).

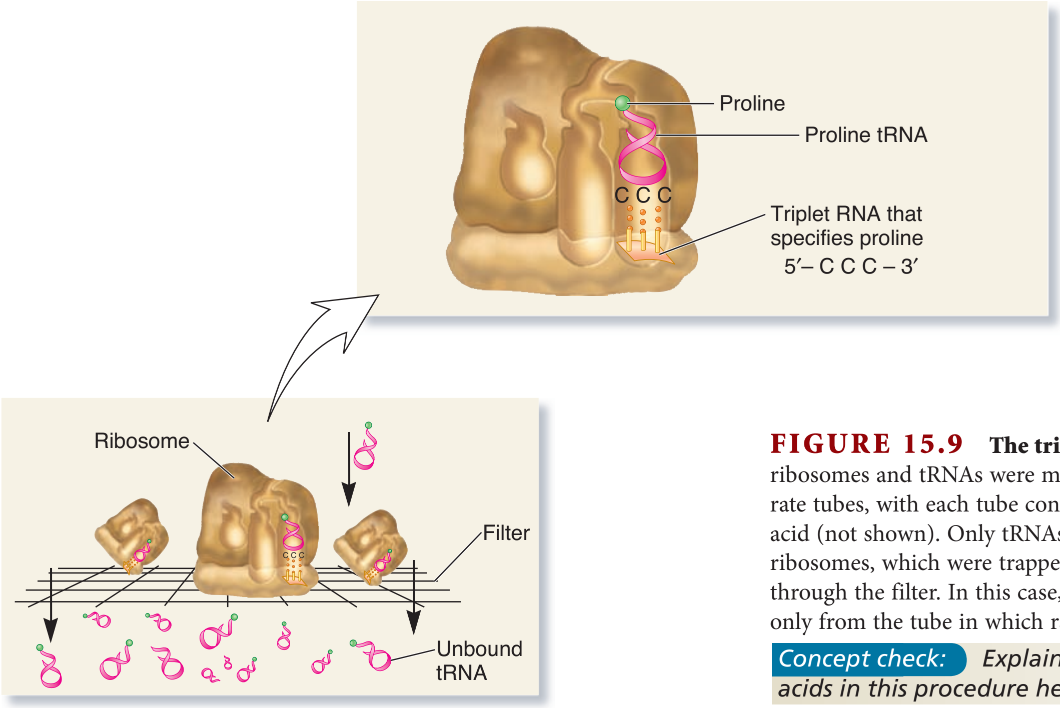


FIGURE 15.9 **The triplet-binding assay.** In this experiment, ribosomes and tRNAs were mixed with a 5′-CCC-3′ triplet in 20 separate tubes, with each tube containing a different radiolabeled amino acid (not shown). Only tRNAs carrying proline became bound to the ribosomes, which were trapped on the filter. Unbound tRNAs passed through the filter. In this case, radioactivity was trapped on the filter only from the tube in which radiolabeled proline was added.

Concept check: *Explain how the use of radiolabeled amino acids in this procedure helped to reveal the genetic code.*

15.3 COMPREHENSION QUESTIONS

1. In the experiments of Crick involving T4 phage, which of the following combination(s) of phage mutants would restore the correct reading frame?
   a. Three different (–) mutations
   b. One (–) mutation and one (+) mutation
   c. Two (–) mutations and one (+) mutation
   d. Both a and b.
2. Let's suppose a researcher mixed together nucleotides with the following percentage of bases: 30% G, 30% C, and 40% A. If RNA was made via polynucleotide phosphorylase, what percentage of the codons would be 5′-GGC-3′?
   a. 30%
   b. 9%
   c. 2.7%
   d. 0%
3. In the triplet binding assay of Nirenberg and Leder, an RNA triplet composed of three bases was able to cause the
   a. translation of a polypeptide.
   b. binding of a tRNA carrying the appropriate amino acid.
   c. termination of translation.
   d. release of the amino acid from the tRNA.

15.4 STRUCTURE AND FUNCTION OF tRNA

Thus far, we have considered the general features of translation and surveyed the structure and functional significance of cellular proteins. The rest of this chapter is devoted to a molecular description of translation as it occurs in living cells. Biochemical studies of protein synthesis and tRNA molecules began in the 1950s. As work progressed toward an understanding of translation, research revealed that different kinds of RNA molecules are involved in the incorporation of amino acids into growing polypeptides. Francis Crick proposed the **adaptor hypothesis.** According to this idea, the position of an amino acid within a polypeptide is determined by the binding between the mRNA and an adaptor molecule carrying a specific amino acid. Later, work by Paul Zamecnik and Mahlon Hoagland suggested that the adaptor molecule is transfer RNA (tRNA). During translation, a tRNA has two functions: (1) It recognizes a three-base codon sequence in mRNA, and (2) it carries an amino acid specific for that codon. In this section, we will examine the structure and function of tRNA molecules.

The Function of a tRNA Depends on the Specificity Between the Amino Acid It Carries and Its Anticodon

The adaptor hypothesis proposes that tRNA molecules recognize the codons within mRNA and carry the correct amino acids to the site of polypeptide synthesis. During mRNA-tRNA recognition, the anticodon in a tRNA molecule binds to a codon in mRNA due to their complementary sequences (**Figure 15.10**). The codon/anticodon binding occurs in an antiparallel manner. Importantly, the anticodon in the tRNA corresponds to the amino acid that it carries. For example, if the anticodon in the tRNA is 3′–AAG–5′, it is complementary to a 5′–UUC–3′ codon. According to the genetic code described earlier in this chapter, the UUC codon specifies phenylalanine. Therefore, the tRNA with a 3′–AAG–5′ anticodon must carry a phenylalanine.

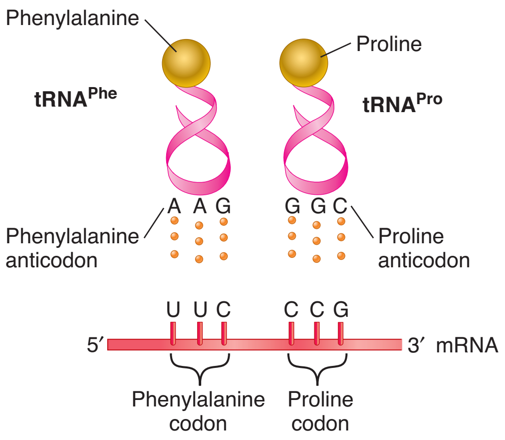


FIGURE 15.10 **Recognition between tRNAs and mRNA.** The anticodon in the tRNA binds to a complementary sequence in the mRNA. At its other end, the tRNA carries the amino acid that corresponds to the codon in the mRNA via the genetic code.

Concept check: *What are the two key functional sites of a tRNA molecule?*

As another example, if the tRNA has a 3′–GGC–5′ anticodon, it is complementary to a 5′–CCG–3′ codon that specifies proline. This tRNA must carry proline.

Recall that the genetic code has 64 codons. Of these, 61 are sense codons that specify the 20 amino acids. Therefore, to synthesize proteins, a cell must produce many different tRNA molecules having specific anticodon sequences. To do so, the chromosomal DNA contains many distinct tRNA genes that encode tRNA molecules with different sequences. According to the adaptor hypothesis, the anticodon in a tRNA specifies the type of amino acid that it carries. Due to this specificity, tRNA molecules are named according to the amino acid they carry. For example, a tRNA that carries phenylalanine is described as tRNA^{Phe}, whereas a tRNA that carries proline is tRNA^{Pro}.

Common Structural Features Are Shared by All tRNAs

To understand how tRNAs act as carriers of the correct amino acids during translation, researchers have examined the structural characteristics of these molecules in great detail. Though a cell makes many different tRNAs, all tRNAs share common structural features. As originally proposed by Robert Holley in 1965, the secondary structure of a tRNA exhibits a cloverleaf pattern. A tRNA has three stem-loops; a few variable sites (locations with additional nucleotides not found in all tRNA molecules); and an acceptor stem with a 3′ single-stranded region (**Figure 15.11**). The acceptor stem is where an amino acid becomes attached to a tRNA (see inset). A conventional numbering system for the nucleotides within a tRNA molecule begins at the 5′ end and proceeds toward the 3′ end. The anticodon is located in the second loop region.

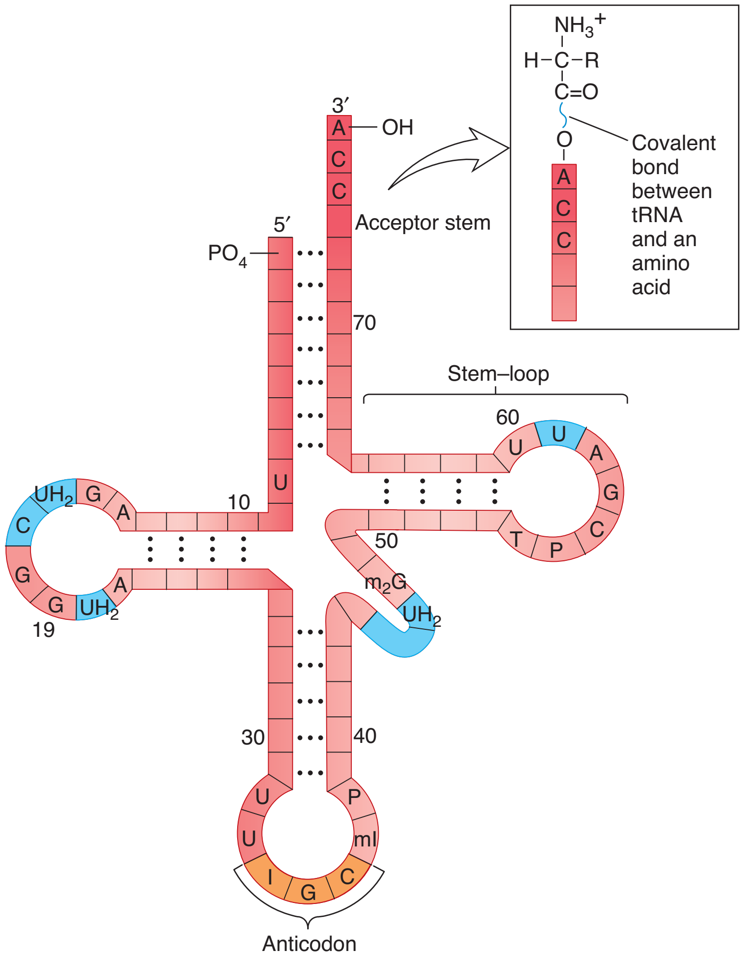


FIGURE 15.11 **Secondary structure of tRNA.** The conventional numbering of nucleotides begins at the 5′ end and proceeds toward the 3′ end. In all tRNAs, the nucleotides at the 3′ end contain the sequence CCA. Certain locations can have additional nucleotides not found in all tRNA molecules. These variable sites are shown in blue. The figure also shows the locations of a few modified bases specifically found in a yeast tRNA that carries alanine. The modified bases are as follows: I = inosine, mI = methylinosine, T = ribothymidine, UH_2 = dihydrouridine, m_2G = dimethylguanosine, and P = pseudouridine. The inset shows an amino acid covalently attached to the 3′ end of a tRNA. Note: The 5′ to 3′ orientation of the anticodon in this drawing (left to right) is opposite to that of tRNAs shown in other drawings, such as Figure 15.10.

The three-dimensional, or tertiary, structure of tRNA molecules involves additional folding of the secondary structure. In the tertiary structure of tRNA, the stem-loops are folded into a much more compact molecule. The ability of RNA molecules to form stem-loops and the tertiary folding of tRNA molecules are described in Chapter 11 (see Figures 11.18 and 11.19). In addition to A, U, G, and C bases, tRNA molecules commonly contain modified bases. For example, Figure 15.11 illustrates a tRNA that contains several modified bases. Among many different species, researchers have found that more than 80 different base modifications can occur in tRNA molecules. We will explore the significance of modified bases in codon recognition later in this chapter.

Aminoacyl-tRNA Synthetases Charge tRNAs by Attaching the Appropriate Amino Acid

To function correctly, each type of tRNA must have the appropriate amino acid attached to its 3′ end. How does an amino acid get attached to a tRNA with the correct anticodon? Enzymes in the cell known as **aminoacyl-tRNA synthetases** catalyze the attachment of amino acids to tRNA molecules. Cells produce 20 different aminoacyl-tRNA synthetase enzymes, 1 for each of the 20 distinct amino acids. Each aminoacyl-tRNA synthetase is named for the specific amino acid it attaches to tRNA. For example, alanyl-tRNA synthetase recognizes a tRNA with an alanine anticodon—tRNA^{Ala}—and attaches an alanine to it.

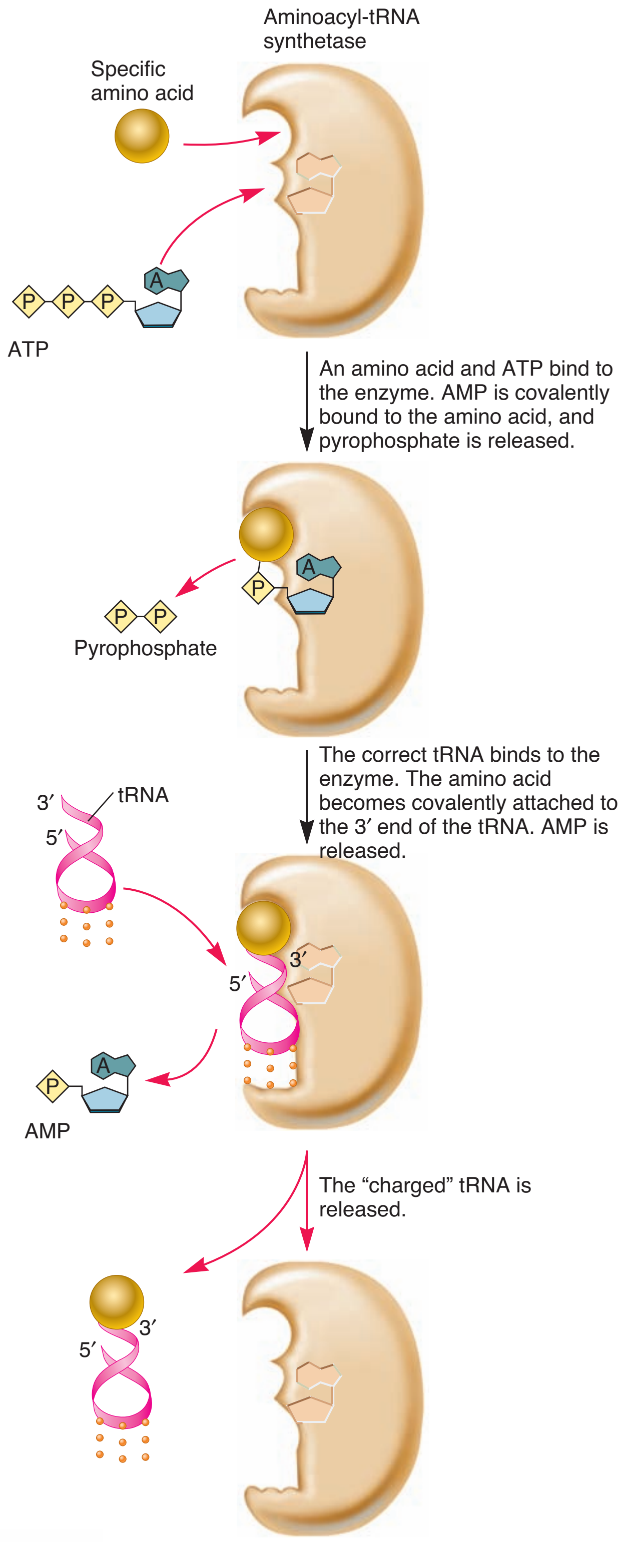
Aminoacyl-tRNA synthetases catalyze a chemical reaction involving three different molecules: an amino acid, a tRNA molecule, and ATP. In the first step of the reaction, a synthetase recognizes a specific amino acid and also ATP (**Figure 15.12**). The ATP is hydrolyzed, resulting in the attachment of AMP to the amino acid and the release of pyrophosphate. During the second step, the correct tRNA binds to the synthetase. The amino acid becomes covalently attached to the 3′ end of the tRNA molecule at the acceptor stem, and AMP is released. Finally, the tRNA with its attached amino acid is released from the enzyme. At this stage, the tRNA is called a **charged tRNA,** or an **aminoacyl-tRNA.** In a charged tRNA molecule, the amino acid is attached to the 3′ end of the tRNA by a covalent bond (see Figure 15.11 inset).

The ability of the aminoacyl-tRNA synthetases to recognize tRNAs has sometimes been called the "second genetic code." This recognition process is necessary to maintain the fidelity of genetic information. The frequency of error for aminoacyl-tRNA synthetases is less than 10^{-5}. In other words, the wrong amino acid is attached to a tRNA less than once in 100,000 times! As you might expect, the anticodon region of the tRNA is usually important for precise recognition by the correct aminoacyl-tRNA synthetase. In studies of *Escherichia coli* synthetases, 17 of the 20 types of aminoacyl-tRNA synthetases recognize the anticodon region of the tRNA. However, other regions of the tRNA are also important recognition sites. These include the acceptor stem and bases in the stem-loops.

As mentioned previously, tRNA molecules frequently contain bases within their structure that have been chemically modified. These modified bases can have important effects on tRNA function. For example, modified bases within tRNA molecules affect the rate of translation and the recognition of tRNAs by aminoacyl-tRNA synthetases. Positions 34 and 37 contain the largest variety of modified nucleotides; position 34 is the first base (in the 5′ to 3′ direction) in the anticodon that matches the third base in the codon of mRNA (also see Figure 15.11). As discussed next, a modified base at position 34 can have important effects on codon-anticodon recognition.

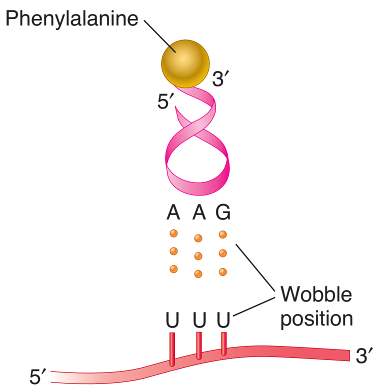
Mismatches That Follow the Wobble Rules Can Occur at the Third Position in Codon-Anticodon Pairing

After considering the structure and function of tRNA molecules, let's reexamine some subtle features of the genetic code.


Aminoacyl-tRNA synthetase
Specific amino acid
P
P
P
A
ATP
An amino acid and ATP bind to the enzyme. AMP is covalently bound to the amino acid, and pyrophosphate is released.
P
P
P
A
Pyrophosphate
The correct tRNA binds to the enzyme. The amino acid becomes covalently attached to the 3′ end of the tRNA. AMP is released.
3′
tRNA
5′
3′
5′
P
A
AMP
The "charged" tRNA is released.
3′
5′

ONLINE ANIMATION **FIGURE 15.12 Catalytic function of aminoacyl-tRNA synthetase.** Aminoacyl-tRNA synthetase has binding sites for a specific amino acid, ATP, and a particular tRNA. In the first step, the enzyme catalyzes the covalent attachment of AMP to an amino acid, yielding an activated amino acid. In the second step, the activated amino acid is attached to the appropriate tRNA.

Concept check: What is the difference between a charged tRNA versus an uncharged tRNA?



(a) Location of wobble position

Wobble position of tRNA anticodon	Third nucleotide of mRNA codon
G	C, U
C	G
A	U, C, G, (A)
U	A, U, G, (C)
I	U, C, A
xm^{5}s^{2}U	A, (G)
xm^{5}Um	
Um	
xm^{5}U	
xo^{5}U	U, A, G
k^{2}C	A

(b) Revised wobble rules

FIGURE 15.13 Wobble position and base-pairing rules. **(a)** The wobble position occurs between the first base (meaning the first base in the 5′ to 3′ direction) in the anticodon and the third base in the mRNA codon. **(b)** The revised wobble rules are slightly different from those originally proposed by Crick. The standard bases found in RNA are G, C, A, and U. In addition, the structures of bases in tRNAs may be modified. Some modified bases that may occur in the wobble position in tRNA are I = inosine; xm^{5}s^{2}U = 5-methyl-2-thiouridine; xm^{5}Um = 5-methyl-2′-*O*-methyluridine; Um = 2′-*O*-methyluridine; xm^{5}U = 5-methyluridine; xo^{5}U = 5-hydroxyuridine; k^{2}C = lysidine (a cytosine derivative). The mRNA bases in parentheses are recognized very poorly by the tRNA.

Concept check: How do the wobble rules affect the total number of different tRNAs that are needed to carry out translation?

As discussed earlier, the genetic code is degenerate, which means that more than one codon can specify the same amino acid. Degeneracy usually occurs at the third position in the codon (**Figure 15.13a**). For example, valine is specified by GUU, GUC, GUA, and GUG. In all four cases, the first two bases are G and U. The third base, however, can be U, C, A, or G. To explain this pattern of degeneracy, Francis Crick proposed in 1966 that it is due to "wobble" at the third position in the codon-anticodon recognition process. According to the **wobble rules,** the first two positions pair strictly according to the AU/GC rule. However, the third position can tolerate certain types of mismatches (see **Figure 15.13b**). This proposal suggested that the base at the third position in the codon does not have to hydrogen bond as precisely with the corresponding base in the anticodon.

Because of the wobble rules, some flexibility is observed in the recognition between a codon and anticodon during the process of translation. When two or more tRNAs that differ at the wobble base are able to recognize the same codon, these are termed **isoacceptor tRNAs.** As an example, tRNAs with an anticodon of 3′–CCA–5′ or 3′–CCG–5′ can recognize a codon with the sequence of 5′–GGU–3′. In addition, the wobble rules enable a single type of tRNA to recognize more than one codon. For example, a tRNA with an anticodon sequence of 3′–AAG–5′ can recognize a 5′–UUC–3′ and a 5′–UUU–3′ codon. The 5′–UUC–3′ codon is a perfect match with this tRNA. The 5′–UUU–3′ codon is mismatched according to the standard RNA-RNA hybridization rules (namely, G in the anticodon is mismatched to U in the codon), but the two can fit according to the wobble rules described in Figure 15.13b. Likewise, the modification of the wobble base to an inosine allows a tRNA to recognize three different codons. At the cellular level, the ability of a single tRNA to recognize more than one codon makes it unnecessary for a cell to make 61 different tRNA molecules with anticodons that are complementary to the 61 possible sense codons. *E. coli* cells, for example, make a population of tRNA molecules that have just 40 different anticodon sequences.

15.4 REVIEWING THE KEY CONCEPTS

- The anticodon in a tRNA is complementary to a codon in mRNA. The tRNA carries a specific amino acid that corresponds to the codon in the mRNA according to the genetic code (see Figure 15.10).
- The secondary structure of tRNA resembles a cloverleaf. The anticodon is in the second loop, and the amino acid is attached to the 3′ end (see Figure 15.11).
- Aminoacyl-tRNA synthetases are a group of enzymes that attach the correct amino acid to a tRNA. The resulting tRNA is called a charged tRNA or an aminoacyl-tRNA (see Figure 15.12).
- Mismatches are allowed between the pairing of tRNAs and mRNA according to the wobble rules (see Figure 15.13).

15.4 COMPREHENSION QUESTIONS

1. If a tRNA has an anticodon with the sequence 5′-CAG-3′, which amino acid does it carry?
   a. Aspartic acid
   b. Leucine
   c. Valine
   d. Glutamine
2. The anticodon of a tRNA is located in the
   a. 3′ single-stranded region of the acceptor stem.
   b. loop of the first stem-loop.
   c. loop of the second stem-loop.
   d. loop of the third stem-loop.
3. An enzyme known as ______________ attaches an amino acid to the ___________ of a tRNA, thereby producing ___________.
   a. aminoacyl-tRNA synthetase, anticodon, a charged tRNA
   b. aminoacyl-tRNA synthetase, 3′ single-stranded region of the acceptor stem, a charged tRNA
   c. polynucleotide phosphorylase, anticodon, a charged tRNA
   d. polynucleotide phosphorylase, anticodon, an aminoacyl tRNA

15.5 RIBOSOME STRUCTURE AND ASSEMBLY

In Section 15.4, we examined how the structure and function of tRNA molecules are important in translation. According to the adaptor hypothesis, tRNAs bind to mRNA due to complementarity between the anticodons and codons. Concurrently, the tRNA molecules have the correct amino acid attached to their 3′ ends.

To synthesize a polypeptide, additional events must occur. In particular, the bond between the 3′ end of the tRNA and the amino acid must be broken, and a peptide bond must be formed between the adjacent amino acids. To facilitate these events, translation occurs on the surface of a macromolecular complex known as the **ribosome.** The ribosome can be thought of as the macromolecular arena where translation takes place.

In this section, we begin by outlining the biochemical compositions of ribosomes in bacterial and eukaryotic cells. We will then examine the key functional sites on ribosomes for the translation process.

Bacterial and Eukaryotic Ribosomes Are Assembled from rRNA and Proteins

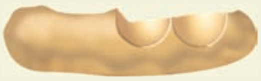
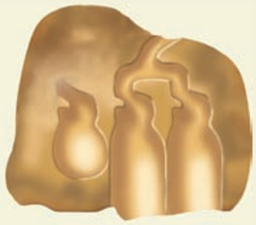
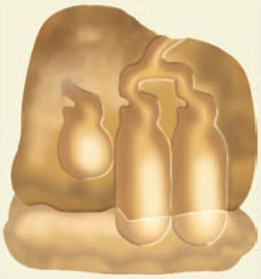
Bacterial cells have one type of ribosome that is found within the cytoplasm. Eukaryotic cells contain biochemically distinct ribosomes in different cellular locations. The most abundant type of ribosome functions in the cytosol, which is the region of the eukaryotic cell that is inside the plasma membrane but outside the membrane-bound organelles. Besides the cytosolic ribosomes, all eukaryotic cells have ribosomes within the mitochondria. In addition, plant cells and algae have ribosomes in their chloroplasts. The compositions of mitochondrial and chloroplast ribosomes are quite different from that of the cytosolic ribosomes. Unless otherwise noted, the term *eukaryotic ribosome* refers to ribosomes in the cytosol, not to those found within organelles. Likewise, the description of eukaryotic translation refers to translation via cytosolic ribosomes.

Each ribosome is composed of structures called the large and small subunits. This term is perhaps misleading because each ribosomal subunit itself is formed from the assembly of many different proteins and RNA molecules called **ribosomal RNA,** or **rRNA.** In bacterial ribosomes, the 30S subunit is formed from the assembly of 21 different ribosomal proteins and a 16S rRNA molecule; the 50S subunit contains 34 different proteins and 5S and 23S rRNA molecules (Table 15.5). The designations 30S and 50S refer to the rate that these subunits sediment when subjected to a centripetal force. This rate is described as a sedimentation coefficient in Svedberg units (S), in honor of Theodor Svedberg, who invented the ultracentrifuge. Together, the 30S and 50S subunits form a 70S ribosome. (Note: Svedberg units do not add up linearly.) In bacteria, the ribosomal proteins and rRNA molecules are synthesized in the cytoplasm, and the ribosomal subunits are assembled there.

The synthesis of eukaryotic rRNA occurs within the nucleus, and the ribosomal proteins are made in the cytosol, where translation takes place. The 40S subunit is composed of 33 proteins and an 18S rRNA; the 60S subunit is made of 49 proteins and 5S, 5.8S, and 28S rRNAs (see Table 15.5). The assembly of the rRNAs and ribosomal proteins to make the 40S and 60S subunits occurs within the **nucleolus,** a region of the nucleus specialized for this purpose. The 40S and 60S subunits are then exported into the cytosol, where they associate to form an 80S ribosome during translation.

TABLE **15.5**

A Comparison of Ribosome Composition in Bacteria and Eukaryotes

	 **Small Subunit**	 **Large Subunit**	 **Assembled Ribosome**
Bacterial			
Sedimentation coefficient	30S	50S	70S
Number of proteins	21	34	55
rRNA	16S rRNA	5S rRNA, 23S rRNA	16S rRNA, 5S rRNA, 23S rRNA
Eukaryotic			
Sedimentation coefficient	40S	60S	80S
Number of proteins	33	49	82
rRNA	18S rRNA	5S rRNA, 5.8S rRNA, 28S rRNA	18S rRNA, 5S rRNA, 5.8S rRNA, 28S rRNA

Components of Ribosomal Subunits Form Functional Sites for Translation

To understand the structure and function of the ribosome at the molecular level, researchers must determine the locations and functional roles of the individual ribosomal proteins and rRNAs. In recent years, many advances have been made toward a molecular understanding of ribosomes. Microscopic and biophysical methods have been used to study ribosome structure. An electron micrograph of bacterial ribosomes is shown in **Figure 15.14a.** More recently, a few research groups have succeeded in crystallizing ribosomal subunits and even intact ribosomes. This is an amazing technical feat, because it is difficult to find the right conditions under which large macromolecules form highly ordered crystals. **Figure 15.14b** shows the crystal structure of bacterial ribosomal subunits. The overall shape of each subunit is largely determined by the structure of the rRNAs, which constitute most of the mass of the ribosome. The interface between the 30S and 50S subunits is primarily composed of rRNA. Ribosomal proteins cluster on the outer surface of the ribosome and on the periphery of the interface.

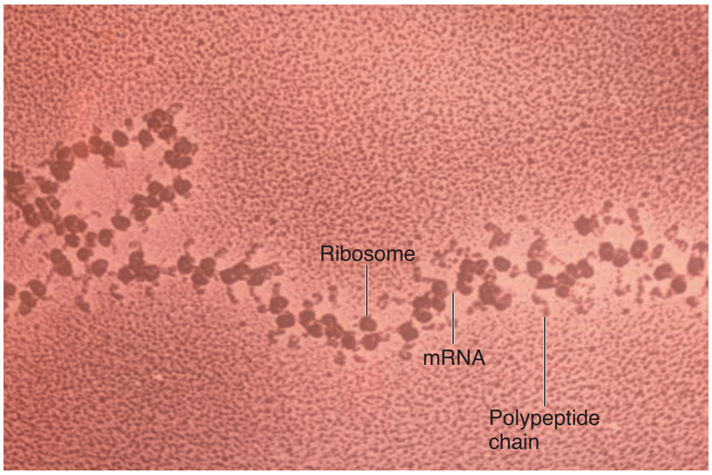
During bacterial translation, the mRNA lies on the surface of the 30S subunit within a space between the 30S and 50S subunits. As the polypeptide is being synthesized, it exits through a channel within the 50S subunit (**Figure 15.14c**). Ribosomes contain discrete sites where tRNAs bind and the polypeptide is synthesized. In 1964, James Watson was the first to propose a two-site model for tRNA binding to the ribosome. These sites are known as the **peptidyl site (P site)** and **aminoacyl site (A site).** In 1981, Knud Nierhaus, Hans Sternbach, and Hans-Jorg Rheinberger proposed a three-site model. This model incorporated the observation that uncharged tRNA molecules can bind to a site on the ribosome that is distinct from the P and A sites. This third site is now known as the **exit site (E site).** The locations of the E, P, and A sites are shown in Figure 15.14c. Next, we will examine the roles of these sites during the stages of translation.

15.5 REVIEWING THE KEY CONCEPTS

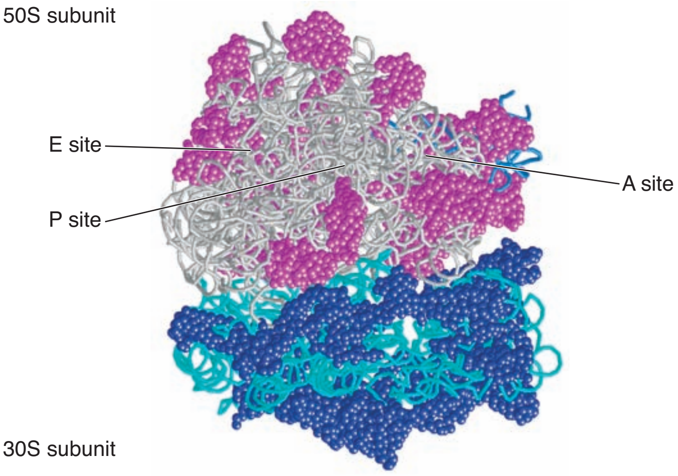
- Ribosomes are the site of polypeptide synthesis. The small and large subunits of ribosomes are composed of rRNAs and multiple proteins (see Table 15.5).
- A ribosome contains an A (aminoacyl), P (peptidyl), and E (exit) site, which are occupied by tRNA molecules (see Figure 15.14).

15.5 COMPREHENSION QUESTIONS

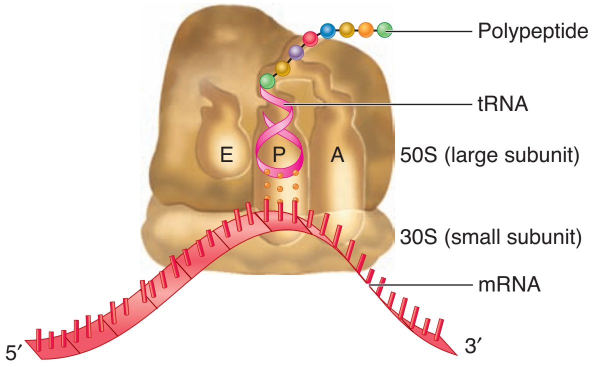
1. Each ribosomal subunit is composed of
   a. multiple proteins.
   b. rRNA.
   c. tRNA.
   d. both a and b.
2. The site(s) on a ribosome where tRNA molecules may be located include the
   a. A site.
   b. P site.
   c. E site.
   d. all of the above.



(a) Ribosomes as seen with electron microscope



(b) Bacterial ribosome model based on X–ray diffraction studies



(c) Model for bacterial ribosome structure

FIGURE 15.14 Ribosomal structure. **(a)** Electron micrograph of ribosomes attached to a bacterial mRNA molecule. **(b)** Crystal structure of the 50S and 30S subunits in bacteria. This model shows the interface between the two subunits. The rRNA is shown in gray strands (50S subunit) and turquoise strands (30S subunit), and proteins are shown in violet (50S subunit) and navy blue (30S subunit). **(c)** A model depicting the sites where tRNA and mRNA bind to an intact ribosome. The mRNA lies on the surface of the 30S subunit. The E, P, and A sites are formed at the interface between the large and small subunits. The growing polypeptide exits through a hole in the 50S subunit.

15.6 STAGES OF TRANSLATION

Like transcription, the process of translation can be viewed as occurring in three stages: initiation, elongation, and termination. **Figure 15.15** presents an overview of these stages. During **initiation,** the ribosomal subunits, mRNA, and the first tRNA assemble to form a complex. After the initiation complex is formed, the ribosome slides along the mRNA in the 5′ to 3′ direction, moving over the codons. This is the **elongation** stage of translation. As the ribosome moves, tRNA molecules sequentially bind to the mRNA in the ribosome, bringing with them the appropriate amino acids. Therefore, amino acids are linked in the order dictated by the codon sequence in the mRNA. Finally, a stop codon is reached, signaling the **termination** of translation. At this point, disassembly occurs, and the newly made polypeptide is released. In this section, we will examine the components required for the translation process and consider their functional roles during the three stages of translation.

The Initiation Stage Involves the Binding of mRNA and the Initiator tRNA to the Ribosomal Subunits

During initiation, an mRNA and the first tRNA bind to the ribosomal subunits. A specific tRNA functions as the **initiator tRNA,** which recognizes the start codon in the mRNA. In bacteria, the initiator tRNA, which is also designated tRNA$^{\mathrm{fMet}}$, carries a methionine that has been covalently modified to *N*-formylmethionine. In this modification, a formyl group (—CHO) is attached to the nitrogen atom in methionine after the methionine has been attached to the tRNA.

Figure 15.16 describes the initiation stage of translation in bacteria during which the mRNA, tRNA$^{\mathrm{fMet}}$, and ribosomal

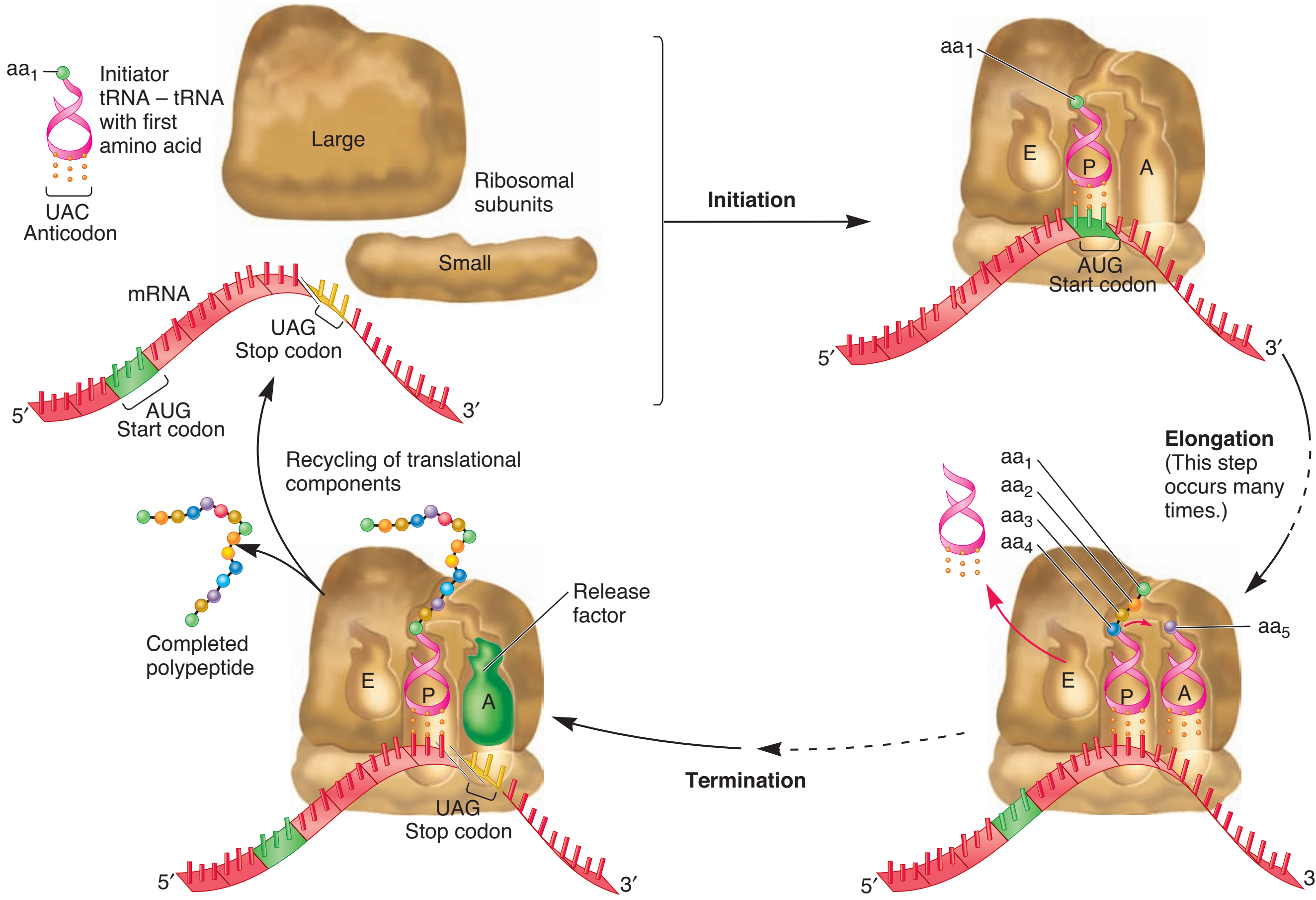

aa1
Initiator tRNA – tRNA with first amino acid
UAC
Anticodon
Large
Ribosomal subunits
Small
mRNA
UAG
Stop codon
5′
AUG
Start codon
3′
Initiation
aa1
E
P
A
AUG
Start codon
5′
3′
Elongation
(This step occurs many times.)
aa1
aa2
aa3
aa4
aa5
E
P
A
5′
3′
Termination
Recycling of translational components
Completed polypeptide
Release factor
E
P
A
UAG
Stop codon
5′
3′

FIGURE 15.15 **Overview of the stages of translation.** The initiation stage involves the assembly of the ribosomal subunits, mRNA, and the initiator tRNA carrying the first amino acid. During elongation, the ribosome slides along the mRNA and synthesizes a polypeptide. Translation ends when a stop codon is reached and the polypeptide is released from the ribosome. (Note: In this and succeeding figures in this chapter, the ribosomes are drawn schematically to emphasize different aspects of the translation process. The structures of ribosomes are described in Figure 15.14.)

Genes → Traits The ability of genes to produce an organism's traits relies on the molecular process of gene expression. During translation, the codon sequence within mRNA (which is derived from a gene sequence during transcription) is translated into a polypeptide sequence. After polypeptides are made within a living cell, they function as components of proteins to govern an organism's traits. For example, once the β-globin polypeptide is made, it functions within the hemoglobin protein and provides red blood cells with the ability to carry oxygen, a vital trait for survival. Translation allows functional proteins to be made within living cells.

Concept check: *Explain how mRNA plays a role in all three stages.*

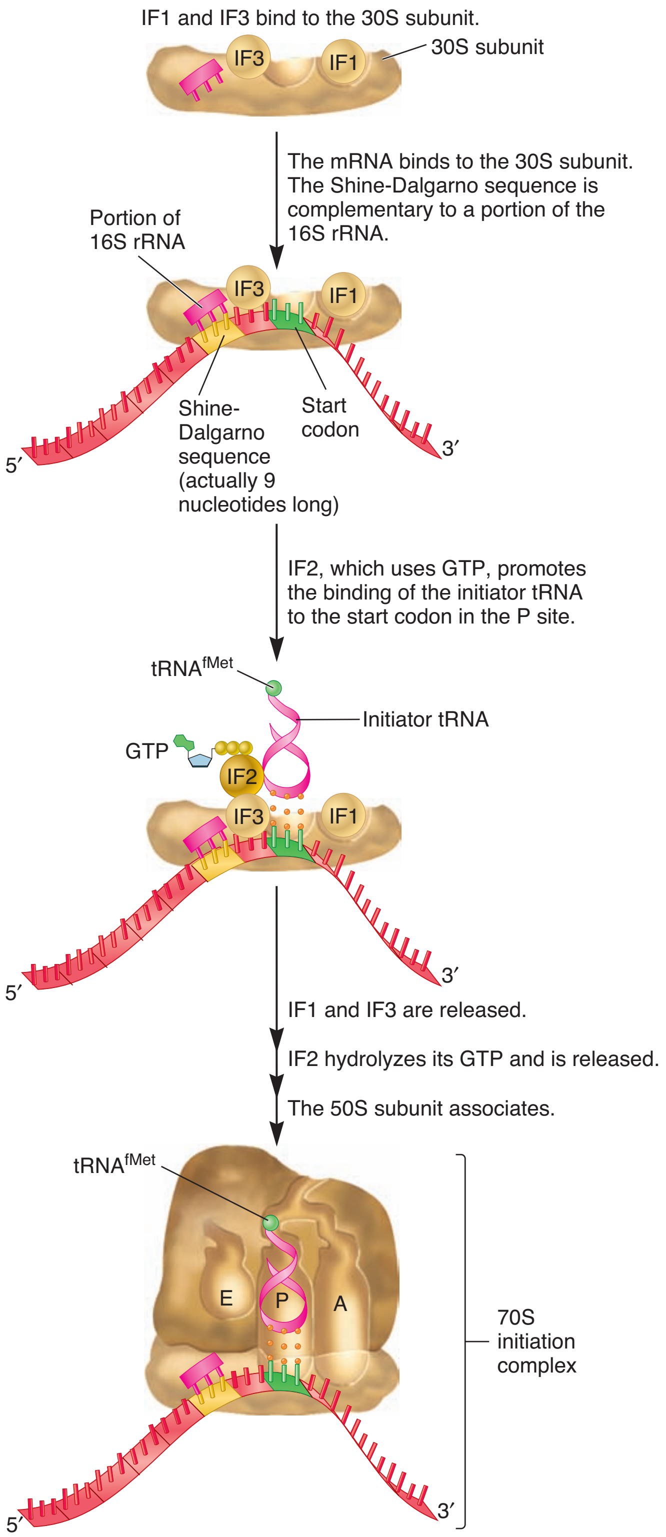
subunits associate with each other to form an initiation complex. The formation of this complex requires the participation of three initiation factors: IF1, IF2, and IF3. First, IF1 and IF3 bind to the 30S subunit. IF1 and IF3 prevent the association of the 50S subunit. Next, the mRNA binds to the 30S subunit. This binding is facilitated by a nine-nucleotide sequence within the bacterial mRNA called the **Shine-Dalgarno sequence.** The location of this sequence is shown in Figure 15.16 and in more detail in **Figure 15.17**. How does the Shine-Dalgarno sequence facilitate the binding of mRNA to the ribosome? The Shine-Dalgarno sequence is complementary to a short sequence within the 16S rRNA, which promotes the hydrogen bonding of the mRNA to the 30S subunit.

Next, tRNA$^{\text{fMet}}$ binds to the mRNA that is already attached to the 30S subunit (see Figure 15.16). This step requires the function of IF2, which uses GTP. The tRNA$^{\text{fMet}}$ binds to the start codon, which is typically a few nucleotides downstream from the Shine-Dalgarno sequence. The start codon is usually AUG, but in some cases it can be GUG or UUG. Even when the start codon is GUG (which normally encodes valine) or UUG (which normally encodes leucine), the first amino acid in the polypeptide is still a formylmethionine because only a tRNA$^{\text{fMet}}$ can initiate translation. During or after translation of the entire polypeptide, the formyl group or the entire formylmethionine may be removed. Therefore, some polypeptides may not have formylmethionine or methionine as their first amino acid. As noted in Figure 15.16, the tRNA$^{\text{fMet}}$ binds to the P site on the ribosome. IF1 is thought to occupy a portion of the A site, thereby preventing the binding of tRNA$^{\text{fMet}}$ to the A site during initiation. By comparison, during the elongation stage that is discussed later, all of the other tRNAs initially bind to the A site.

After the mRNA and tRNA$^{\text{fMet}}$ have become bound to the 30S subunit, IF1 and IF3 are released, and then IF2 hydrolyzes its GTP and is also released. This allows the 50S ribosomal subunit to associate with the 30S subunit. Much later, after translation is completed, IF1 binding is necessary to dissociate the 50S and 30S ribosomal subunits so that the 30S subunit can reinitiate with another mRNA molecule.

In eukaryotes, the assembly of the initiation complex bears similarities to that in bacteria. However, as described in **Table 15.6**, additional factors are required for the initiation process. Note that the initiation factors are designated eIF (for eukaryotic Initiation Factor) to distinguish them from bacterial initiation factors. The initiator tRNA in eukaryotes carries methionine rather than formylmethionine, as in bacteria. A eukaryotic initiation factor, eIF2, binds directly to tRNA$^{\text{Met}}$ to recruit it to the 40S subunit. Eukaryotic mRNAs do not have a Shine-Dalgarno sequence. How then are eukaryotic mRNAs recognized by the ribosome? The mRNA is recognized by eIF4, which is a multiprotein complex that recognizes the 7-methylguanosine cap and facilitates the binding of the mRNA to the 40S subunit.

The identification of the correct AUG start codon in eukaryotes occurs in a different manner compared to bacteria. After the initial binding of mRNA to the ribosome, the next step is to locate an AUG start codon that is somewhere downstream from the 5′ cap structure. In 1986, Marilyn Kozak proposed that the ribosome begins at the 5′ end and then scans along the mRNA in the 3′ direction in search of an AUG start codon. In many, but not all cases, the ribosome uses the first AUG codon that it encounters as a start codon. When a start codon is identified, the 60S subunit assembles onto the 40S subunit with the aid of eIF5, forming the initiation complex.


IF1 and IF3 bind to the 30S subunit.
IF3
IF1
30S subunit
The mRNA binds to the 30S subunit. The Shine-Dalgarno sequence is complementary to a portion of the 16S rRNA.
Portion of 16S rRNA
IF3
IF1
Shine-Dalgarno sequence (actually 9 nucleotides long)
Start codon
5′
3′
IF2, which uses GTP, promotes the binding of the initiator tRNA to the start codon in the P site.
tRNA^fMet
Initiator tRNA
GTP
IF2
IF3
IF1
5′
3′
IF1 and IF3 are released.
IF2 hydrolyzes its GTP and is released.
The 50S subunit associates.
tRNA^fMet
E
P
A
70S initiation complex
5′
3′

ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 15.16 **The initiation stage of translation in bacteria.**

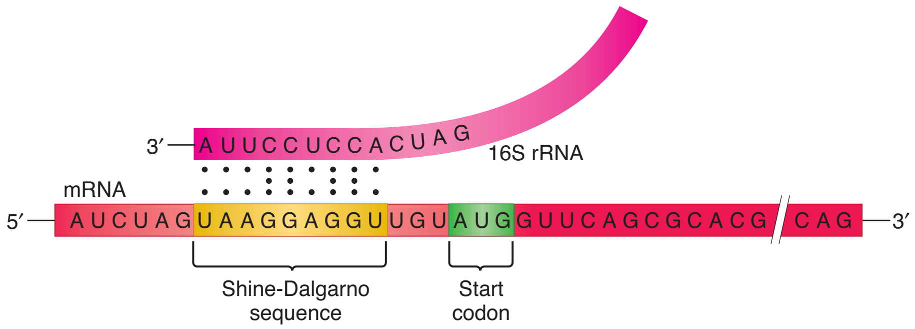

3′
A U U C C U C C A C U A G
16S rRNA
mRNA
5′
A U C U A G U A A G G A G G U U G U A U G G U U C A G C G C A C G
C A G
3′
Shine-Dalgarno sequence
Start codon

FIGURE 15.17 The locations of the Shine-Dalgarno sequence and the start codon in bacterial mRNA. The Shine-Dalgarno sequence is complementary to a sequence in the 16S rRNA. It hydrogen bonds with the 16S rRNA to promote initiation. The start codon is typically a few nucleotides downstream from the Shine-Dalgarno sequence.

Concept check: Why does a bacterial mRNA bind specifically to the small ribosomal subunit?

By analyzing the sequences of many eukaryotic mRNAs, researchers have found that not all AUG codons near the 5′ end of mRNA can function as start codons. In some cases, the scanning ribosome passes over the first AUG codon and chooses an AUG farther down the mRNA. The sequence of bases around the AUG codon plays an important role in determining whether or not it is selected as the start codon by a scanning ribosome. The consensus sequence for optimal start codon recognition in more complex eukaryotes, such as vertebrates and vascular plants, is shown here.

						Start	*Codon*		
G	C	C	(A/G)	C	C	A	U	G	G
−6	−5	−4	−3	−2	−1	+1	+2	+3	+4

Aside from an AUG codon itself, a guanine at the +4 position and a purine, preferably an adenine, at the −3 position are the most important sites for start codon selection. These rules for optimal translation initiation are called **Kozak's rules.**

Polypeptide Synthesis Occurs During the Elongation Stage

During the elongation stage of translation, amino acids are added, one at a time, to a growing polypeptide (**Figure 15.18**). Even though this process is rather complex, it occurs at a remarkable rate. Under normal cellular conditions, a polypeptide can elongate at a rate of 15 to 20 amino acids per second in bacteria and 2 to 6 amino acids per second in eukaryotes!

To begin elongation, a charged tRNA brings a new amino acid to the ribosome so it can be attached to the end of the growing polypeptide. At the top of Figure 15.18, a short polypeptide is attached to the tRNA located at the P site of the ribosome. A charged tRNA carrying a single amino acid binds to the A site. This binding occurs because the anticodon in the tRNA is complementary to the codon in the mRNA. The hydrolysis of GTP by the elongation factor, EF-Tu, provides energy for the binding of a tRNA to the A site. In addition, the 16S rRNA, which is a component of the small 30S ribosomal subunit, plays a key role that ensures the proper recognition between the mRNA

TABLE **15.6**

A Simplified Comparison of Translational Protein Factors in Bacteria and Eukaryotes

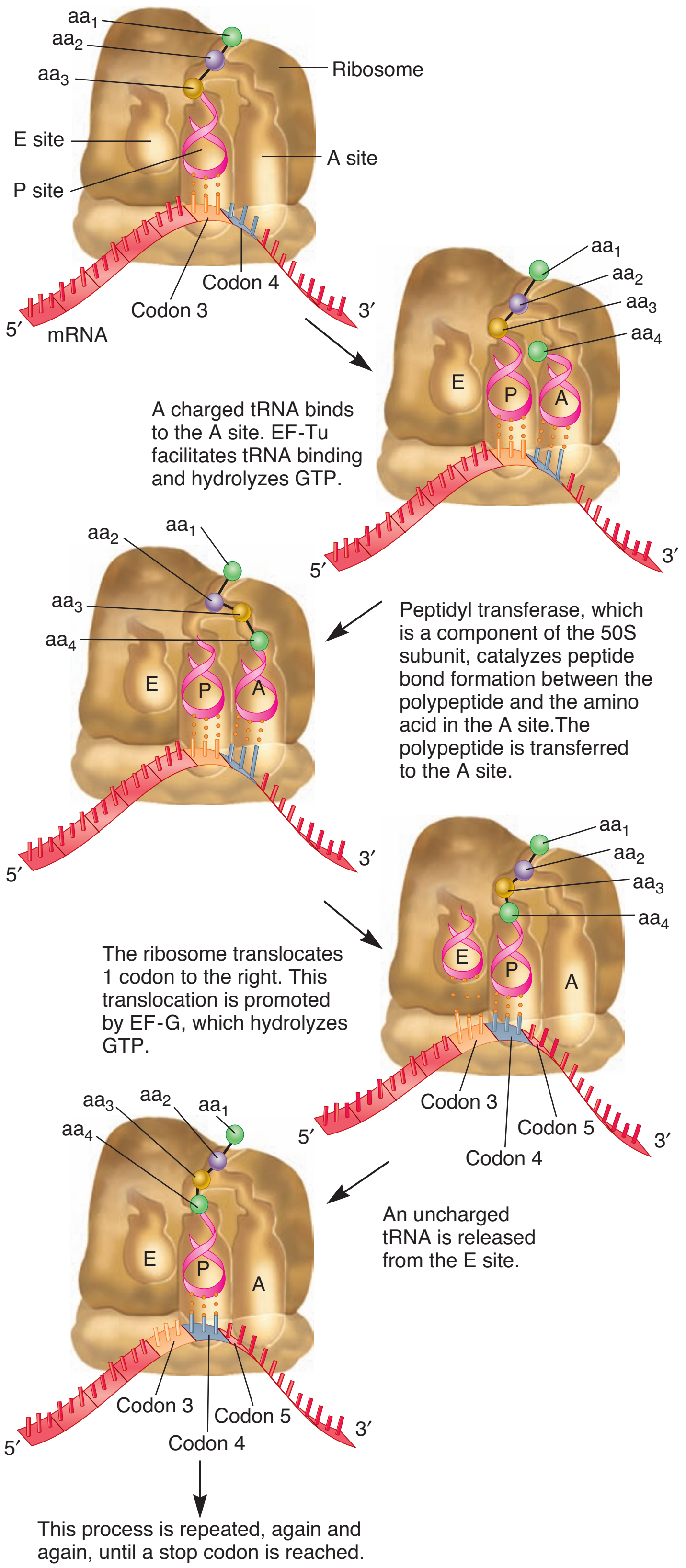
Bacterial Factors	Eukaryotic Factors*	Function
Initiation Factors		
	eIF4	Involved with the recognition of the 7-methylguanosine cap and the binding of the mRNA to the small ribosomal subunit
IF1, IF3	eIF1, eIF3, eIF6	Prevent the association between the small and large ribosomal subunits and favor their dissociation
IF2	eIF2	Promotes the binding of the initiator tRNA to the small ribosomal subunit
	eIF5	Helps to dissociate the other elongation factors, which allows the large ribosomal subunit to bind
Elongation Factors		
EF-Tu	eEF1α	Involved in the binding of tRNAs to the A site
EF-Ts	eEF1βγ	Nucleotide exchange factors required for the functioning of EF-Tu and eEF1α, respectively
EF-G	eEF2	Required for translocation
Release Factors		
RF1, RF2	eRF1	Recognize a stop codon and trigger the cleavage of the polypeptide from the tRNA
RF3	eRF3	GTPases that are also involved in termination

*Eukaryotic translation factors are typically composed of multiple proteins.

and correct tRNA. The 16S rRNA can detect when an incorrect tRNA is bound at the A site and prevents elongation until the mispaired tRNA is released from the A site. This phenomenon, termed the **decoding function** of the ribosome, is important in maintaining a high fidelity of mRNA translation. An incorrect amino acid is incorporated into a growing polypeptide at a rate of approximately one mistake per 10,000 amino acids, or 10^{-4}.

The next step of elongation is the **peptidyl transfer** reaction—the polypeptide is removed from the tRNA in the P site and transferred to the amino acid at the A site. This transfer is accompanied by the formation of a peptide bond between the amino acid at the A site and the polypeptide, lengthening the polypeptide by one amino acid. The peptidyl transfer reaction is catalyzed by a component of the 50S subunit known as **peptidyl transferase,** which is composed of several proteins and rRNA. In 2000, based on the crystal structure of the 50S subunit, Thomas Steitz, Peter Moore, and their colleagues concluded that the 23S rRNA (not the ribosomal protein) catalyzes bond formation between adjacent amino acids. In other words, the ribosome is a ribozyme!

After the peptidyl transfer reaction is complete, the ribosome moves, or translocates, to the next codon in the mRNA. This moves the tRNAs at the P and A sites to the E and P sites, respectively. Finally, the uncharged tRNA exits the E site. You should notice that the next codon in the mRNA is now exposed in the unoccupied A site. At this point, a charged tRNA can enter the empty A site, and the same series of steps can add the next amino acid to the growing polypeptide. As you may have realized, the A, P, and E sites are named for the role of the tRNA that is usually found there. The A site binds an aminoacyl-tRNA (also called a charged tRNA), the P site usually contains the peptidyl-tRNA (a tRNA with an attached peptide), and the E site is where the uncharged tRNA exits.



A charged tRNA binds to the A site. EF-Tu facilitates tRNA binding and hydrolyzes GTP.

Peptidyl transferase, which is a component of the 50S subunit, catalyzes peptide bond formation between the polypeptide and the amino acid in the A site.The polypeptide is transferred to the A site.

The ribosome translocates 1 codon to the right. This translocation is promoted by EF-G, which hydrolyzes GTP.

An uncharged tRNA is released from the E site.

This process is repeated, again and again, until a stop codon is reached.

ONLINE ANIMATION INTERACTIVE EXERCISE

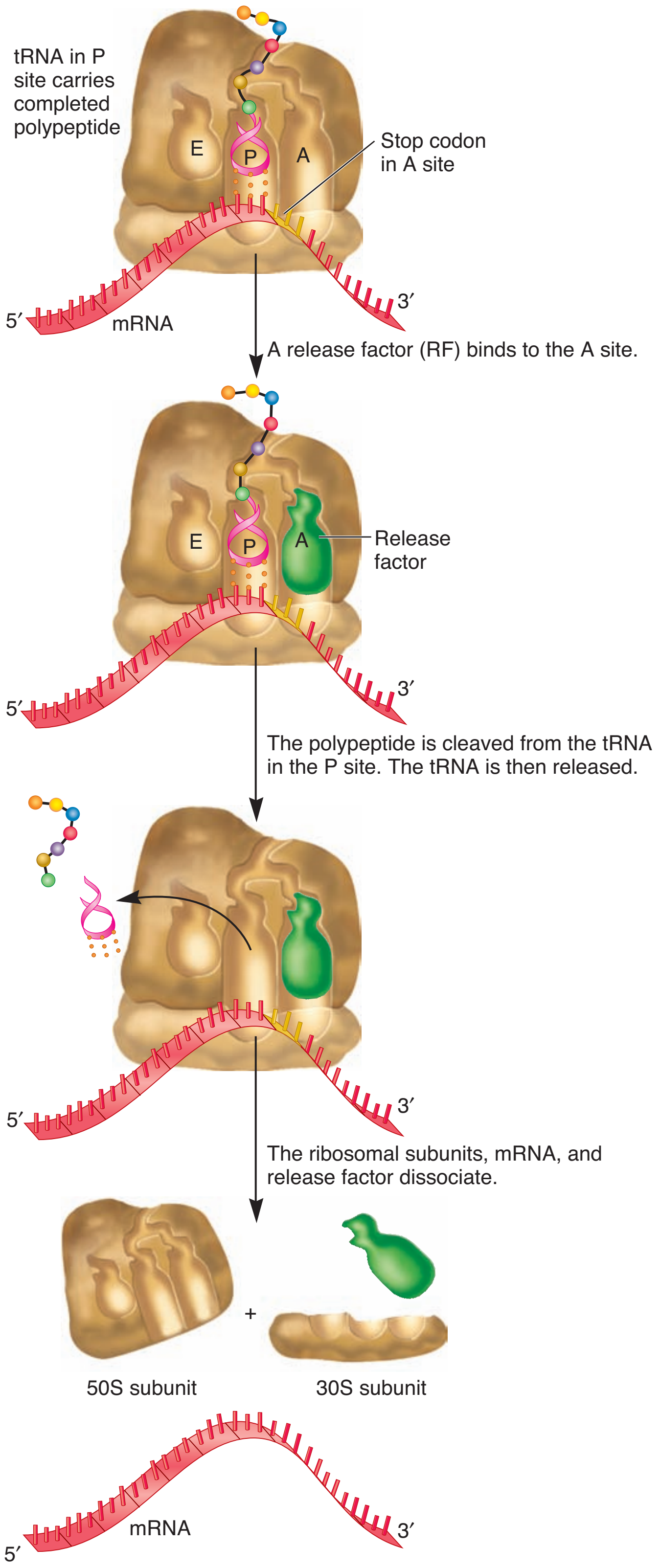
FIGURE 15.18 The elongation stage of translation in bacteria. This process begins with the binding of an incoming tRNA. The hydrolysis of GTP by EF-Tu provides the energy for the binding of the tRNA to the A site. A peptide bond is then formed between the incoming amino acid and the last amino acid in the growing polypeptide. This moves the polypeptide to the A site. The ribosome then translocates in the 3′ direction so the two tRNAs are moved to the E and P sites. The tRNA carrying the polypeptide is now back in the P site. This translocation requires the hydrolysis of GTP via EF-G. The uncharged tRNA in the E site is released from the ribosome. Now the process is ready to begin again. Each cycle of elongation causes the polypeptide to grow by one amino acid.

Concept check: What is the role of peptidyl transferase during the elongation stage?

Termination Occurs When a Stop Codon Is Reached in the mRNA

The final stage of translation, known as termination, occurs when a stop codon is reached in the mRNA. In most species, the three stop codons are UAA, UAG, and UGA. The stop codons are not recognized by a tRNA with a complementary sequence. Instead, they are recognized by proteins known as **release factors** (see Table 15.6). Interestingly, the three-dimensional structures of release factor proteins are "molecular mimics" that resemble the structure of tRNAs. Such proteins can specifically bind to a stop codon sequence. In bacteria, RF1 recognizes UAA and UAG, and RF2 recognizes UGA and UAA. A third release factor, RF3, is also required. In eukaryotes, a single release factor, eRF1, recognizes all three stop codons, and eRF3 is also required for termination.

Figure 15.19 illustrates the termination stage of translation in bacteria. At the top of this figure, the completed polypeptide is attached to a tRNA in the P site. A stop codon is located at the A site. In the first step, RF1 or RF2 binds to the stop codon at the A site and RF3 (not shown) binds at a different location on the ribosome. After RF1 (or RF2) and RF3 have bound, the bond between the polypeptide and the tRNA is hydrolyzed. The polypeptide and tRNA are then released from the ribosome. The final step in translational termination is the disassembly of ribosomal subunits, mRNA, and the release factors.


tRNA in P site carries completed polypeptide
E
P
A
Stop codon in A site
5′
mRNA
3′
A release factor (RF) binds to the A site.
E
P
A
Release factor
5′
3′
The polypeptide is cleaved from the tRNA in the P site. The tRNA is then released.
5′
3′
The ribosomal subunits, mRNA, and release factor dissociate.
+
50S subunit
30S subunit
mRNA
5′
3′

ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 15.19 The termination stage of translation in bacteria. When a stop codon is reached, RF1 or RF2 binds to the A site. (RF3 binds elsewhere and uses GTP to facilitate the termination process.) The polypeptide is cleaved from the tRNA in the P site and released. The tRNA is released, and the rest of the components disassemble.

Concept check: Explain why release factors are called "molecular mimics."

Bacterial Translation Can Begin Before Transcription Is Completed

Although most of our knowledge concerning transcription and translation has come from genetic and biochemical studies, electron microscopy (EM) has also been an important tool in elucidating the mechanisms of transcription and translation. As described earlier in this chapter, EM has been a critical technique in facilitating our understanding of ribosome structure. In addition, it has been employed to visualize genetic processes such as translation.

The first success in the EM observation of gene expression was achieved by Oscar Miller, Jr., and his colleagues in 1967. Figure 15.20 shows an electron micrograph of a bacterial gene in the act of gene expression. Prior to this experiment, biochemical and genetic studies had suggested that the translation of a bacterial structural gene begins before the mRNA transcript is completed. In other words, as soon as an mRNA strand is long enough, a ribosome attaches to the 5′ end and begins translation, even before RNA polymerase has reached the transcriptional termination site within the gene. This phenomenon is termed the "coupling between transcription and translation" in bacterial cells. Note that coupling of these processes does not usually occur in eukaryotes, because transcription takes place in the nucleus, whereas translation occurs in the cytosol.

As shown in Figure 15.20, several RNA polymerase enzymes have recognized a gene and begun to transcribe it. Because the transcripts on the right side are longer than those on the left, Miller concluded that transcription was proceeding from left to right in the micrograph. This EM image also shows

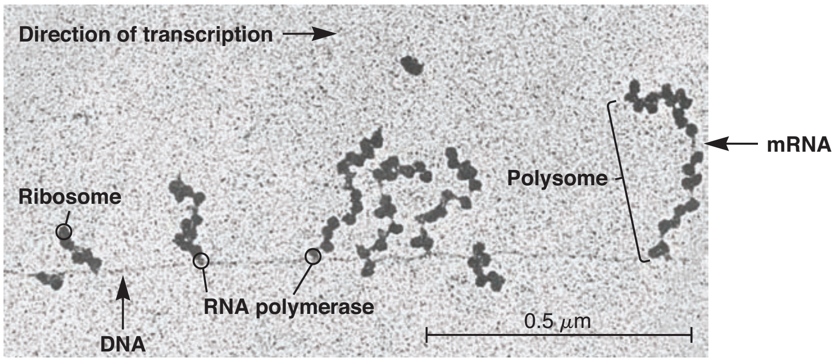


FIGURE 15.20 **Coupling between transcription and translation in bacteria.** An electron micrograph showing the simultaneous transcription and translation processes. The DNA is transcribed by many RNA polymerases that move along the DNA from left to right. Note that the RNA transcripts are getting longer as you go from left to right. Ribosomes attach to the mRNA, even before transcription is completed. The complex of many ribosomes bound to the same mRNA is called a polyribosome, or a polysome. Several polyribosomes are seen here.

the process of translation. Relatively small mRNA transcripts, near the left side of the figure, have a few ribosomes attached to them. As the transcripts become longer, additional ribosomes are attached to them. The term **polyribosome,** or **polysome,** is used to describe an mRNA transcript that has many bound ribosomes in the act of translation. In this electron micrograph, the growing polypeptides were too small for researchers to observe. In later studies, as EM techniques became more refined, the polypeptides emerging from the ribosome were also visible (see Figure 15.14a).

Bacterial and Eukaryotic Translation Show Similarities and Differences

Throughout this chapter, we have compared translation in both bacteria and eukaryotic organisms. The general steps of translation are similar in all forms of life, but we have also seen some striking differences between bacteria and eukaryotes. Table 15.7 compares translation between these groups.

15.6 REVIEWING THE KEY CONCEPTS

- The three stages of translation are initiation, elongation, and termination (see Figure 15.15).
- During the initiation stage of translation, the mRNA, initiator tRNA, and ribosomal subunits assemble. Initiation factors are involved in the process. In bacteria, the Shine-Dalgarno sequence promotes the binding of the mRNA to the small ribosomal subunit (see Figures 15.16, 15.17, Table 15.6).
- During elongation, tRNAs bring amino acids to the A site and a series of peptidyl transferase reactions creates a polypeptide. At each step, the polypeptide is transferred from the P site to the A site. The tRNAs are released from the E site. Elongation factors are involved in this process (see Figure 15.18).
- Start codon selection in complex eukaryotes follows Kozak's rules.
- During termination, a release factor binds to a stop codon in the A site. This promotes the cleavage of the polypeptide from the tRNA and the subsequent disassembly of the tRNA, mRNA, and ribosomal subunits (see Figure 15.19).

TABLE **15.7**

Comparison of Bacterial and Eukaryotic Translation

	Bacterial	Eukaryotic
Ribosome composition:	70S ribosomes: 30S subunit— 21 proteins + 1 rRNA 50S subunit— 34 proteins + 2 rRNAs	80S ribosomes: 40S subunit— 33 proteins + 1 rRNA 60S subunit— 49 proteins + 3 rRNAs
Initiator tRNA:	$\text{tRNA}^{\text{fMet}}$	tRNA^{Met}
Formation of the initiation complex:	Requires IF1, IF2, and IF3	Requires more initiation factors compared to bacterial initiation
Initial binding of mRNA to the ribosome:	Requires a Shine-Dalgarno sequence	Requires a 7-methylguanosine cap
Selection of a start codon:	AUG, GUG, or UUG located just downstream from the Shine-Dalgarno sequence	According to Kozak's rules
Elongation rate	Typically 15 to 20 amino acids per second	Typically 2 to 6 amino acids per second
Termination:	Requires RF1, RF2, and RF3	Requires eRF1 and eRF3
Coupled to transcription:	Yes	No

- Bacterial translation can begin before transcription is completed (see Figure 15.20).
- Bacterial and eukaryotic translation show many similarities and differences (see Table 15.7).

15.6 COMPREHENSION QUESTIONS

1. During the initiation stage of translation in bacteria, which of the following event(s) occurs?
   a. IF1 and IF3 bind to the 30S subunit.
   b. The mRNA binds to the 30S subunit and $\text{tRNA}^{\text{fMet}}$ binds to the start codon in the mRNA.
   c. IF2 hydrolyzes its GTP and is released; the 50S subunit binds to the 30S subunit.
   d. All of the above events occur.
2. The Kozak rules determine
   a. the choice of the start codon in complex eukaryotes.
   b. the choice of the start codon in bacteria.
   c. the site in the mRNA where translation ends.
   d. how fast the mRNA is translated.
3. During the peptidyl transferase reaction, the polypeptide, which is attached to a tRNA in the __________ becomes bound via ________ to an amino acid attached to a tRNA in the __________.
   a. A site, several hydrogen bonds, P site
   b. A site, a peptide bond, P site
   c. P site, a peptide bond, A site
   d. P site, several hydrogen bonds, A site
4. A release factor is referred to as a "molecular mimic" because its structure is similar to
   a. a ribosome.
   b. an mRNA.
   c. a tRNA.
   d. an elongation factor.

KEY TERMS

Page 324. translation, structural genes, messenger RNA (mRNA)
Page 325. alkaptonuria, inborn error of metabolism
Page 326. one-gene/one-enzyme hypothesis
Page 327. polypeptide, protein, genetic code, sense codons, start codon, stop codons, termination codons, nonsense codons, transfer RNA (tRNA), anticodons, degenerate
Page 328. synonymous codons, wobble base, reading frame
Page 329. selenocysteine, pyrrolysine, peptide bond, amino-terminal end, N-terminus, carboxyl-terminal end, C-terminus, side chain, R group
Page 330. primary structure, chaperones, secondary structure, α helix, β sheet
Page 331. tertiary structure
Page 332. quaternary structure, subunits
Page 333. frameshift mutation
Page 335. cell-free translation system
Page 338. adaptor hypothesis
Page 340. aminoacyl-tRNA synthetases, charged tRNA, aminoacyl-tRNA
Page 341. wobble rules, isoacceptor tRNAs
Page 342. ribosome, ribosomal RNA (rRNA), nucleolus
Page 343. peptidyl site (P site), aminoacyl site (A site), exit site (E site)
Page 344. initiation, elongation, termination, initiator tRNA
Page 345. Shine-Dalgarno sequence
Page 346. Kozak's rules
Page 347. decoding function, peptidyl transfer, peptidyl transferase
Page 348. release factors
Page 349. polyribosome, polysome

CHAPTER SUMMARY

- Cellular proteins are made via the translation of mRNA.

15.1 The Genetic Basis for Protein Synthesis

- Garrod studied the disease called alkaptonuria and suggested that some genes encode enzymes (see Figure 15.1).
- Beadle and Tatum studied *Neurospora* mutants that were altered in their nutritional requirements and hypothesized that one gene encodes one enzyme. This hypothesis was later modified because: (1) some proteins are not enzymes; (2) some proteins are composed of two or more different polypeptides; and (3) some genes encode RNAs that are not translated into polypeptides (see Figure 15.2).

15.2 The Relationship between the Genetic Code and Protein Synthesis

- During translation, the sequence of codons in mRNA is used via tRNA molecules to make a polypeptide with a specific amino acid sequence (see Figure 15.3).
- The genetic code refers to the relationship between three-base codons in the mRNA and the amino acids that are incorporated into a polypeptide. One codon (AUG) is a start codon, which determines the reading frame of the mRNA. Three codons (UAA, UAG, and UGA) function as stop codons (see Table 15.1).
- The genetic code is largely universal, but some exceptions are known to occur (see Table 15.2).
- A polypeptide is made by the formation of peptide bonds between adjacent amino acids. Each polypeptide has a directionality from its amino-terminal end to its carboxyl-terminal end, which parallels the arrangement of codons in mRNA in the 5′ to 3′ direction (see Figure 15.4).
- Amino acids differ in their side chain structure (see Figure 15.5).
- Protein structure can be viewed at different levels, which consist of primary structure (sequence of amino acids), secondary structure (repeating folding patterns such as the α helix and the β sheet), tertiary structure (additional folding), and

quaternary structure (the binding of multiple subunits to each other) (see Figures 15.6, 15.7).

15.3 Experimental Determination of the Genetic Code

- Crick studied mutations in T4 phage and determined that the genetic code is read in triplets (see Table 15.3).
- Nirenberg and colleagues used synthetic RNA and a cell-free translation system to decipher the genetic code (see Figure 15.8).
- Other methods of deciphering the genetic code included the synthesis of copolymers by Khorana and the triplet-binding assay of Nirenberg and Leder (see Table 15.4, Figure 15.9).

15.4 Structure and Function of tRNA

- The anticodon in a tRNA is complementary to a codon in mRNA. The tRNA carries a specific amino acid that corresponds to the codon in the mRNA according to the genetic code (see Figure 15.10).
- The secondary structure of tRNA resembles a cloverleaf. The anticodon is in the second loop, and the amino acid is attached to the 3′ end (see Figure 15.11).
- Aminoacyl-tRNA synthetases are a group of enzymes that attach the correct amino acid to a tRNA. The resulting tRNA is called a charged tRNA or an aminoacyl-tRNA (see Figure 15.12).
- Mismatches are allowed between the pairing of tRNAs and mRNA according to the wobble rules (see Figure 15.13).

15.5 Ribosome Structure and Assembly

- Ribosomes are the site of polypeptide synthesis. The small and large subunits of ribosomes are composed of rRNAs and multiple proteins (see Table 15.5).
- A ribosome contains an A (aminoacyl), P (peptidyl), and E (exit) site, which are occupied by tRNA molecules (see Figure 15.14).

15.6 Stages of Translation

- The three stages of translation are initiation, elongation, and termination (see Figure 15.15).
- During the initiation stage of translation, the mRNA, initiator tRNA, and ribosomal subunits assemble. Initiation factors are involved in the process. In bacteria, the Shine-Dalgarno sequence promotes the binding of the mRNA to the small ribosomal subunit (see Figures 15.16, 15.17, Table 15.6).
- During elongation, tRNAs bring amino acids to the A site and a series of peptidyl transferase reactions creates a polypeptide. At each step, the polypeptide is transferred from the P site to the A site. The tRNAs are released from the E site. Elongation factors are involved in this process (see Figure 15.18).
- Start codon selection in complex eukaryotes follows Kozak's rules.
- During termination, a release factor binds to a stop codon in the A site. This promotes the cleavage of the polypeptide from the tRNA and the subsequent disassembly of the tRNA, mRNA, and ribosomal subunits (see Figure 15.19).
- Bacterial translation can begin before transcription is completed (see Figure 15.20).
- Bacterial and eukaryotic translation show many similarities and differences (see Table 15.7).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. The first amino acid in a purified bacterial protein is methionine. The start codon in the mRNA is GUG, which codes for valine. Why isn't the first amino acid formylmethionine or valine?

Answer: During polypeptide synthesis, the first amino acid carried by the initiator tRNA is formylmethionine. This occurs even when the start codon is GUG (valine) or UUG (leucine). The formyl group can be later removed to yield methionine as the first amino acid.

S2. A tRNA has the anticodon sequence 3′–CAG–5′. What amino acid does it carry?

Answer: Because the anticodon is 3′–CAG–5′, it would be complementary to a codon with the sequence 5′–GUC–3′. According to the genetic code, this codon specifies the amino acid valine. Therefore, this tRNA must carry valine at its acceptor stem.

S3. In eukaryotic cells, the assembly of ribosomal subunits occurs in the nucleolus. As discussed in Chapter 14 (see Figure 14.16), a single 45S rRNA transcript is cleaved to produce the three rRNA fragments—18S, 5.8S, and 28S rRNA—that play a key role in creating the structure of the ribosome. The genes that encode the 45S precursor are found in multiple copies (i.e., they are moderately repetitive). The segments of chromosomes that contain the 45S rRNA genes align themselves at the center of the nucleolus. This site is called the nucleolar-organizing center. In this region, active transcription of the 45S gene takes place. Briefly explain how the assembly of the ribosomal subunits occurs.

Answer: In the nucleolar-organizing center, the 45S RNA is cleaved to the 18S, 5.8S, and 28s rRNAs. The other components of the ribosomal subunits, 5S rRNA and ribosomal proteins, must also be imported into the nucleolar region. Because proteins are made in the cytosol, they must enter the nucleus through the nuclear pores. When all components are present, they assemble into 40S and 60S ribosomal subunits. Following assembly, the ribosomal subunits exit the nucleus through the nuclear pores and enter the cytosol.

S4. An antibiotic is a drug that kills or inhibits the growth of microorganisms. The use of antibiotics has been of great importance in the battle against many infectious diseases caused by microorganisms. The mode of action for many antibiotics is to inhibit the translation process within bacterial cells. Certain antibiotics selectively bind to bacterial (70S) ribosomes but do not inhibit eukaryotic (80S) ribosomes. Their ability to inhibit translation can occur at different steps in the translation process. For example, tetracycline prevents the attachment of tRNA to the ribosome, whereas erythromycin inhibits the translocation of the ribosome along the mRNA. Why would an antibiotic bind to a bacterial ribosome but not to a eukaryotic ribosome? How does inhibition of translation by antibiotics such as tetracycline prevent bacterial growth?

Answer: Because bacterial ribosomes have a different protein and rRNA composition than eukaryotic ribosomes, certain drugs can recognize these different components and bind specifically to bacterial ribosomes, thereby interfering with the process of translation. In other words, the surface of a bacterial ribosome must be somewhat different from the surface of a eukaryotic ribosome so that the drugs bind to the surface of only bacterial ribosomes. If a bacterial cell is exposed to tetracycline or other antibiotics, it cannot synthesize new polypeptides. Because polypeptides form functional proteins needed for processes such as cell division, the bacterium is unable to grow and proliferate.

Conceptual Questions

C1. An mRNA has the following sequence:

5´–GGCGAUGGGCAAUAAACCGGGCCAGUAAGC–3´

Identify the start codon and determine the complete amino acid sequence that would be translated from this mRNA.

C2. What does it mean when we say that the genetic code is degenerate? Discuss the universality of the genetic code.

C3. According to the adaptor hypothesis, are the following statements true or false?

A. The sequence of an anticodon in a tRNA directly recognizes a codon sequence in mRNA, with some room for wobble.

B. The amino acid attached to the tRNA directly recognizes a codon sequence in mRNA.

C. The amino acid attached to the tRNA affects the binding of the tRNA to a codon sequence in mRNA.

C4. Researchers have isolated strains of bacteria that carry mutations within tRNA genes. These mutations can change the sequence of the anticodon. For example, a normal tRNA^{Trp} gene encodes a tRNA with the anticodon 3′–ACC–5′. A mutation could change this sequence to 3′–CCC–5′. When this mutation occurs, the tRNA still carries a tryptophan at its 3′ acceptor stem, even though the anticodon sequence has been altered.

A. How would this mutation affect the synthesis of polypeptides within the bacterium?

B. What does this mutation tell you about the recognition between tryptophanyl-tRNA synthetase and tRNA^{Trp}? Does the enzyme primarily recognize the anticodon or not?

C5. The covalent attachment of an amino acid to a tRNA is an endergonic reaction. In other words, it requires an input of energy for the reaction to proceed. Where does the energy come from to attach amino acids to tRNA molecules?

C6. The wobble rules for tRNA-mRNA pairing are shown in Figure 15.13b. If we assume that the tRNAs do not contain modified bases, what is the minimum number of tRNAs needed to efficiently recognize the codons for the following types of amino acids?

A. Leucine

B. Methionine

C. Serine

C7. How many different sequences of mRNA could encode a peptide with the sequence proline-glycine-methionine-serine?

C8. If a tRNA molecule carries a glutamic acid, what are the two possible anticodon sequences that it could contain? Be specific about the 5′ and 3′ ends.

C9. A tRNA has an anticodon sequence 3′–GGU–5′. What amino acid does it carry?

C10. If a tRNA has an anticodon sequence 3′–CCI–5′, what codon(s) can it recognize?

C11. Describe the anticodon of a single tRNA that could recognize the codons 5′–AAC–3′ and 5′–AAU–3′. How would this tRNA need to be modified for it to also recognize 5′–AAA–3′?

C12. Describe the structural features that all tRNA molecules have in common.

C13. In the tertiary structure of tRNA, where is the anticodon region relative to the attachment site for the amino acid? Are they adjacent to each other?

C14. What is the role of aminoacyl-tRNA synthetase? The ability of aminoacyl-tRNA synthetases to recognize tRNAs has sometimes been called the "second genetic code." Why has the function of this type of enzyme been described this way?

C15. What is an activated amino acid?

C16. Discuss the significance of modified bases within tRNA molecules.

C17. How and when does formylmethionine become attached to the initiator tRNA in bacteria?

C18. Is it necessary for a cell to make 61 different tRNA molecules, corresponding to the 61 codons for amino acids? Explain your answer.

C19. List the components required for translation. Describe the relative sizes of these different components. In other words, which components are small molecules, macromolecules, or assemblies of macromolecules?

C20. Describe the components of eukaryotic ribosomal subunits and where the assembly of the subunits occurs within living cells.

C21. The term *subunit* can be used in a variety of ways. Compare the use of the term subunit in proteins versus ribosomal subunit.

C22. Do the following events during bacterial translation occur primarily within the 30S subunit, within the 50S subunit, or at the interface between these two ribosomal subunits?

A. mRNA-tRNA recognition

B. Peptidyl transfer reaction

C. Exit of the polypeptide from the ribosome

D. Binding of initiation factors IF1, IF2, and IF3

C23. What are the three stages of translation? Discuss the main events that occur during these three stages.

C24. Describe the sequence in bacterial mRNA that promotes recognition by the 30S subunit.

C25. For each of the following initiation factors, how would eukaryotic initiation of translation be affected if it were missing?

A. eIF2

B. eIF4

C. eIF5

C26. How does a eukaryotic ribosome select its start codon? Describe the sequences in eukaryotic mRNA that provide an optimal context for a start codon.

C27. Rank each of the following sequences in order (from best to worst) as sequences that could be used to initiate translation according to Kozak's rules.

GACGCCAUGG
GCCUCCAUGC
GCCAUCAAGG
GCCACCAUGG

C28. Explain the functional roles of the A, P, and E sites during translation.

C29. An mRNA has the following sequence: 5′–AUG UAC UAU GGG GCG UAA–3′. Describe the amino acid sequence of the polypeptide that would be encoded by this mRNA. Be specific about the amino-terminal and carboxyl-terminal ends.

C30. Which steps during the translation of bacterial mRNA involve an interaction between complementary strands of RNA?

C31. What is the function of the nucleolus?

C32. In which of the ribosomal sites, the A site, P site, and/or E site, could the following be found?

A. A tRNA without an amino acid attached

B. A tRNA with a polypeptide attached

C. A tRNA with a single amino acid attached

C33. What is a polysome?

C34. According to Figure 15.18, explain why the ribosome translocates along the mRNA in a 5′ to 3′ direction rather than a 3′ to 5′ direction.

C35. Lactose permease of *E. coli* is a protein composed of a single polypeptide that is 417 amino acids in length. By convention, the amino acids within a polypeptide are numbered from the amino-terminal end to the carboxyl-terminal end. Are the following questions about lactose permease true or false?

A. Because the 64th amino acid is glycine and the 68th amino acid is aspartic acid, the codon for glycine, 64, is closer to the 3′ end of the mRNA than the codon for aspartic acid, 68.

B. The mRNA that encodes lactose permease must be greater than 1241 nucleotides in length.

C36. An mRNA encodes a polypeptide that is 312 amino acids in length. The 53rd codon in this polypeptide is a tryptophan codon. A mutation in the gene that encodes this polypeptide changes this tryptophan codon into a stop codon. How many amino acids occur in the resulting polypeptide: 52, 53, 259, or 260?

C37. Explain what is meant by the "coupling of transcription and translation." Does coupling occur in bacterial and/or eukaryotic cells? Explain.

Experimental Questions

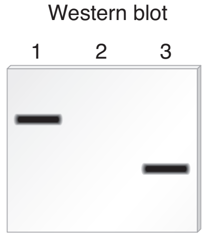
E1. In the experiment of Figure 15.8, what would be the predicted amounts of amino acids incorporated into polypeptides if the RNA was a random polymer containing 50% C and 50% G?

E2. Polypeptides can be translated in vitro. Would a bacterial mRNA be translated in vitro by eukaryotic ribosomes? Would a eukaryotic mRNA be translated in vitro by bacterial ribosomes? Why or why not?

E3. Discuss how the elucidation of the structure of the ribosome can help us to understand its function.

E4. Figure 15.20 shows an electron micrograph of a bacterial gene as it is being transcribed and translated. In this figure, label the 5′ and 3′ ends of the DNA and RNA strands. Place an arrow where you think the start codons are found in the mRNA transcripts.

E5. Chapter 19 describes a blotting method known as Western blotting, which can be used to detect the production of a polypeptide that is translated from a particular mRNA. In this method, a protein is detected with an antibody that specifically recognizes and binds to its amino acid sequence. The antibody acts as a probe to detect the presence of the protein. In a Western blotting experiment, a mixture of cellular proteins is separated using gel electrophoresis according to their molecular masses. After the antibody has bound to the protein of interest within a blot of a gel, the protein is visualized as a dark band. For example, an antibody that recognizes the β-globin polypeptide could be used to specifically detect the β-globin polypeptide in a blot. As shown here, the method of Western blotting can be used to determine the amount and relative size of a particular protein that is produced in a given cell type.



Lane 1 is a sample of proteins isolated from normal red blood cells.

Lane 2 is a sample of proteins isolated from kidney cells. Kidney cells do not produce β globin.

Lane 3 is a sample of proteins isolated from red blood cells from a patient with β-thalassemia. This patient is homozygous for a mutation that results in the shortening of the β-globin polypeptide.

Now here is the question. A protein called troponin contains 334 amino acids. Because each amino acid weighs 120 daltons (Da) (on average), the molecular mass of this protein is about 40,000 Da, or 40 kDa. Troponin functions in muscle cells, and it is not expressed in nerve cells. Draw the expected results of a Western blot for the following samples:

Lane 1: Proteins isolated from muscle cells

Lane 2: Proteins isolated from nerve cells

Lane 3: Proteins isolated from the muscle cells of an individual who is homozygous for a mutation that introduces a stop codon at codon 177

E6. The technique of Western blotting can be used to detect proteins that are translated from a particular mRNA. This method is described in Chapter 19 and also in experimental question E5. Let's suppose a researcher was interested in the effects of mutations on the expression of a structural gene that encodes a protein we will call protein X. This protein is expressed in skin cells and contains 572 amino acids. Its molecular mass is approximately 68,600 Da, or 68.6 kDa. Make a drawing that shows the expected results of a Western blot using proteins isolated from the skin cells obtained from the following individuals:

Lane 1: A normal individual

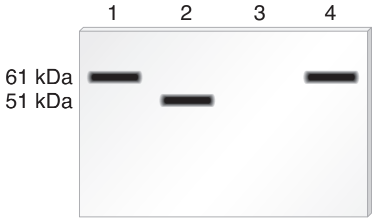
Lane 2: An individual who is homozygous for a deletion, which removes the promoter for this gene

Lane 3: An individual who is heterozygous in which one gene is normal and the other gene has a mutation that introduces an early stop codon at codon 421

Lane 4: An individual who is homozygous for a mutation that introduces an early stop codon at codon 421

Lane 5: An individual who is homozygous for a mutation that changes codon 198 from a valine codon into a leucine codon

E7. The protein known as tyrosinase is needed to make certain types of pigments. Tyrosinase is composed of a single polypeptide with 511 amino acids. Because each amino acid weighs 120 Da (on average), the molecular mass of this protein is approximately 61,300 Da, or 61.3 kDa. People who carry two defective copies of the tyrosinase gene have the condition known as albinism. They are unable to make pigment in the skin, eyes, and hair. Western blotting can be used to detect proteins that are translated from a particular mRNA. This method is described in Chapter 19 and also in experimental question E5. Skin samples were collected from a pigmented individual (lane 1) and from three unrelated albino individuals (lanes 2, 3, and 4) and subjected to a Western blot analysis using an antibody that recognizes tyrosinase.



Explain the possible cause of albinism in the three albino individuals.

E8. Although 61 codons specify the 20 amino acids, most species display a codon bias. This means that certain codons are used much more frequently than other codons. For example, UUA, UUG, CUU, CUC, CUA, and CUG all specify leucine. In yeast, however, the UUG codon is used to specify leucine approximately 80% of the time.

A. The experiment of Figure 15.8 shows the use of a cell-free translation system. In this experiment, the RNA, which was used for translation, was chemically synthesized. Instead of using a chemically synthesized RNA, researchers can isolate mRNA from living cells and then add the mRNA to the cell-free translation system. If a researcher isolated mRNA from kangaroo cells and then added it to a cell-free translation system that came from yeast cells, how might the phenomenon of codon bias affect the production of proteins?

B. Discuss potential advantages and disadvantages of codon bias for translation.

Questions for Student Discussion/Collaboration

1. Discuss why you think the ribosome needs to contain so many proteins and rRNA molecules. Does it seem like a waste of cellular energy to make such a large structure so that translation can occur?
2. Discuss and make a list of the similarities and differences in the events that occur during the initiation, elongation, and termination stages of transcription (see Chapter 14) and translation (Chapter 15).
3. Which events during translation involve molecular recognition between base sequences within different RNAs? Which events involve recognition between different protein molecules?

Answers to Comprehension Questions

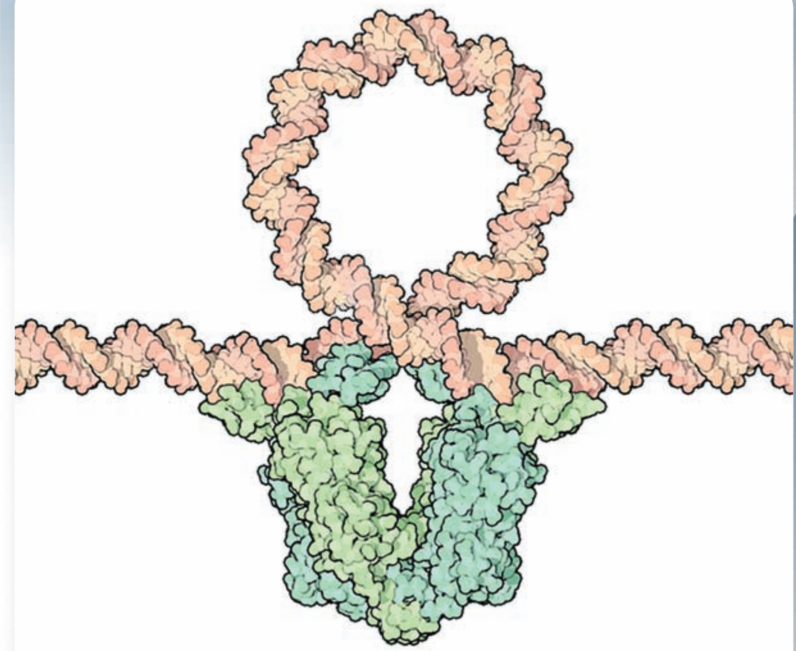
15.1: a, c
15.2: a, d, b, c
15.3: d, c, b
15.4: b, c, b
15.5: d, d
15.6: d, a, c, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

16



A model showing the binding of a genetic regulatory protein to DNA, which results in a DNA loop. *The model shown here involves the lac repressor protein found in E. coli binding to the operator site in the lac operon.*

CHAPTER OUTLINE

16.1 Overview of Transcriptional Regulation

16.2 Regulation of the *lac* Operon

16.3 Regulation of the *trp* Operon

16.4 Translational and Posttranslational Regulation

16.5 Riboswitches

GENE REGULATION IN BACTERIA

The expression of structural genes, which encode polypeptides, ultimately leads to the production of functional proteins. As we discussed in Chapters 14 and 15, gene expression is a multistep process that proceeds from transcription to translation, and it may involve posttranslational effects on protein structure and function. **Gene regulation** is the phenomenon in which the level of gene expression can vary under different conditions. In comparison, unregulated genes have essentially constant levels of expression in all conditions over time. Unregulated genes are also called **constitutive genes.** Frequently, constitutive genes encode proteins that are continuously needed for the survival of the bacterium. In contrast, the majority of genes are regulated so that the proteins they encode can be produced at the proper times and in the proper amounts.

A key benefit of gene regulation is that the encoded proteins are produced only when they are required. Therefore, the cell avoids wasting valuable energy making proteins it does not need. From the viewpoint of natural selection, this enables an organism such as a bacterium to compete as efficiently as possible for limited resources. Gene regulation is particularly important because bacteria exist in an environment that is frequently changing with regard to temperature, nutrients, and many other factors.

The following are a few common processes regulated at the genetic level:

1. ***Metabolism:*** Some proteins function in the metabolism of small molecules. For example, certain enzymes are needed for a bacterium to metabolize particular sugars. These enzymes are required only when the bacterium is exposed to such sugars in its environment.
2. ***Response to environmental stress:*** Certain proteins help a bacterium to survive environmental stress such as osmotic shock or heat shock. These proteins are required only when the bacterium is confronted with the stress.
3. ***Cell division:*** Some proteins are needed for cell division. These are necessary only when the bacterial cell is getting ready to divide.

In this chapter, we will examine the molecular mechanisms that account for these types of gene regulation. As shown in **Figure 16.1**, gene regulation in bacteria can occur at any step in the pathway of gene expression. As you will learn, regulation allows bacteria to conserve their resources by expressing proteins only when they are needed.

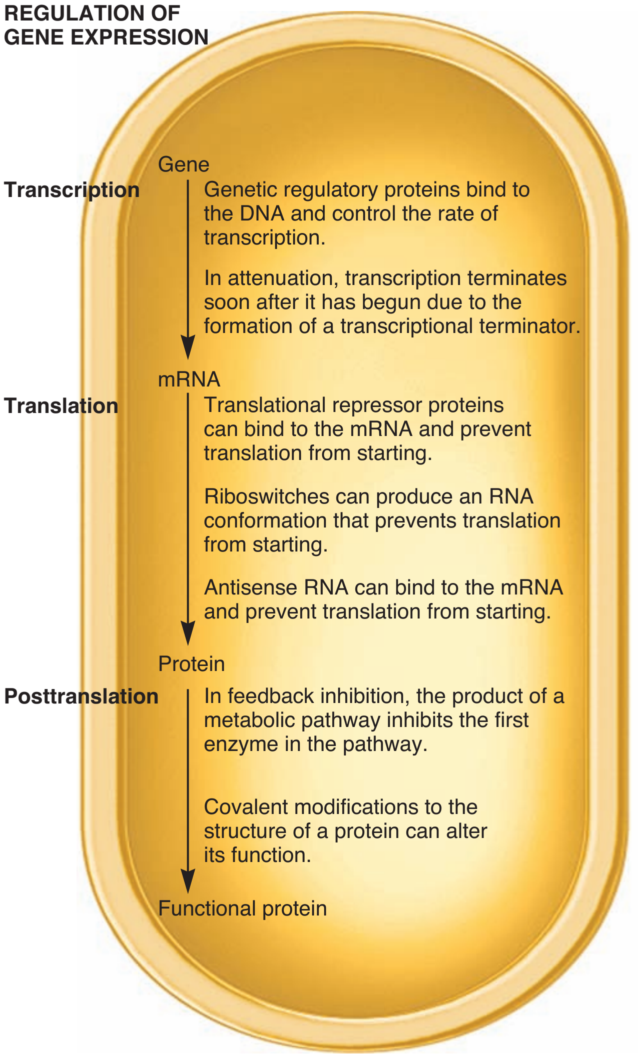

REGULATION OF GENE EXPRESSION
Gene
Transcription
Genetic regulatory proteins bind to the DNA and control the rate of transcription.
In attenuation, transcription terminates soon after it has begun due to the formation of a transcriptional terminator.
mRNA
Translation
Translational repressor proteins can bind to the mRNA and prevent translation from starting.
Riboswitches can produce an RNA conformation that prevents translation from starting.
Antisense RNA can bind to the mRNA and prevent translation from starting.
Protein
Posttranslation
In feedback inhibition, the product of a metabolic pathway inhibits the first enzyme in the pathway.
Covalent modifications to the structure of a protein can alter its function.
Functional protein

FIGURE 16.1 **Common points where regulation of gene expression occurs in bacteria.**

Concept check: *What is an advantage of gene regulation?*

16.1 OVERVIEW OF TRANSCRIPTIONAL REGULATION

In bacteria, the most common way to regulate gene expression is by influencing the rate at which transcription is initiated. Although we frequently refer to genes as being "turned on or off," it is more accurate to say that the level of gene expression is increased or decreased. At the level of transcription, this means that the rate of RNA synthesis can be increased or decreased.

In most cases, transcriptional regulation involves the actions of regulatory proteins that can bind to the DNA and affect the rate of transcription of one or more nearby genes. Two types of regulatory proteins are common. A **repressor** is a regulatory protein that binds to the DNA and inhibits transcription, whereas an **activator** is a regulatory protein that increases the rate of transcription. Transcriptional regulation by a repressor protein is termed **negative control,** and regulation by an activator protein is considered to be **positive control.**

In conjunction with regulatory proteins, small effector molecules often play a critical role in transcriptional regulation. However, small effector molecules do not bind directly to the DNA to alter transcription. Rather, an effector molecule exerts its effects by binding to an activator or repressor. The binding of the effector molecule causes a conformational change in the regulatory protein, usually influencing whether or not the protein can bind to the DNA. Genetic regulatory proteins that respond to small effector molecules have two functional domains. One domain is a site where the protein binds to the DNA; the other domain is the binding site for the effector molecule.

Regulatory proteins are given names describing how they affect transcription when they are bound to the DNA (repressor or activator). In contrast, small effector molecules are given names that describe how they affect transcription when they are present in the cell at a sufficient concentration to exert their effect (Figure 16.2). An **inducer** is a small effector molecule that causes transcription to increase. An inducer may accomplish this in two ways: It could bind to a repressor protein and prevent it from binding to the DNA, or it could bind to an activator protein and cause it to bind to the DNA. In either case, the transcription rate is increased. Genes that are regulated in this manner are called **inducible genes.**

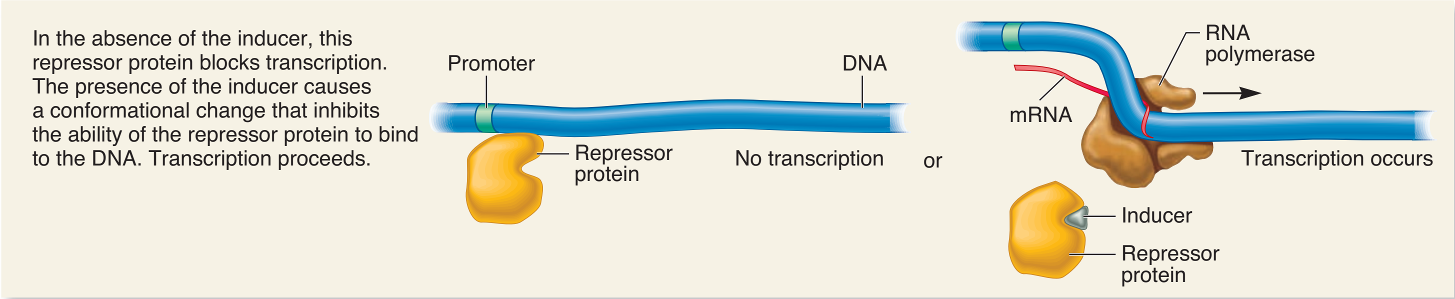
Alternatively, the presence of a small effector molecule may inhibit transcription. This can also occur in two ways. A **corepressor** is a small molecule that binds to a repressor protein, thereby causing the protein to bind to the DNA. An **inhibitor** binds to an activator protein and prevents it from binding to the DNA. Both corepressors and inhibitors act to reduce the rate of transcription. Therefore, the genes they regulate are termed **repressible genes.** Unfortunately, this terminology can be confusing because a repressible system could involve an activator protein, or an inducible system could involve a repressor protein.

16.1 REVIEWING THE KEY CONCEPTS

- Repressors and activators are proteins that exert negative control or positive control, respectively. Small effector molecules control the function of repressors and activators. These include inducers, inhibitors, and corepressors (see Figure 16.2).

16.1 COMPREHENSION QUESTIONS

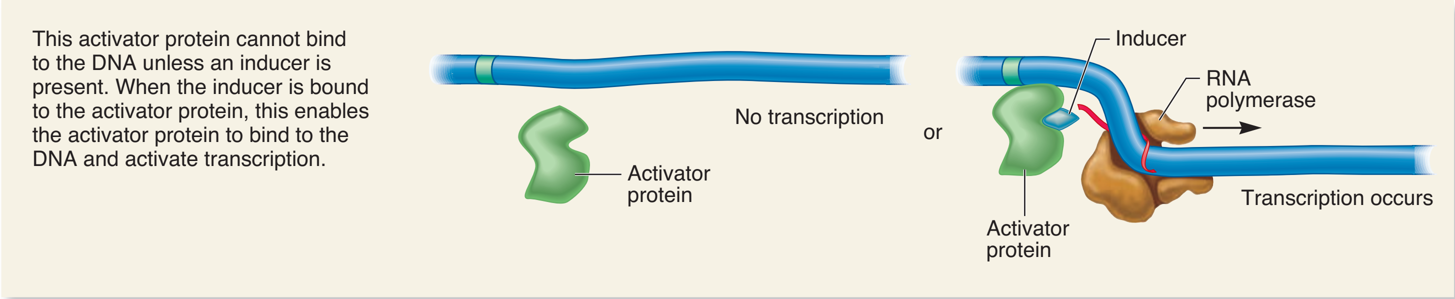
1. A repressor is a __________ that __________ transcription.
   a. small effector molecule, inhibits
   b. small effector molecule, enhances
   c. regulatory protein, inhibits
   d. regulatory protein, enhances.
2. Which of the following conditions would cause transcription to be activated?
   a. A repressor plus an inducer
   b. A repressor plus a corepressor
   c. An activator plus an inhibitor
   d. None of the above



In the absence of the inducer, this repressor protein blocks transcription. The presence of the inducer causes a conformational change that inhibits the ability of the repressor protein to bind to the DNA. Transcription proceeds.

Promoter
DNA
Repressor protein
No transcription
or
RNA polymerase
mRNA
Transcription occurs
Inducer
Repressor protein

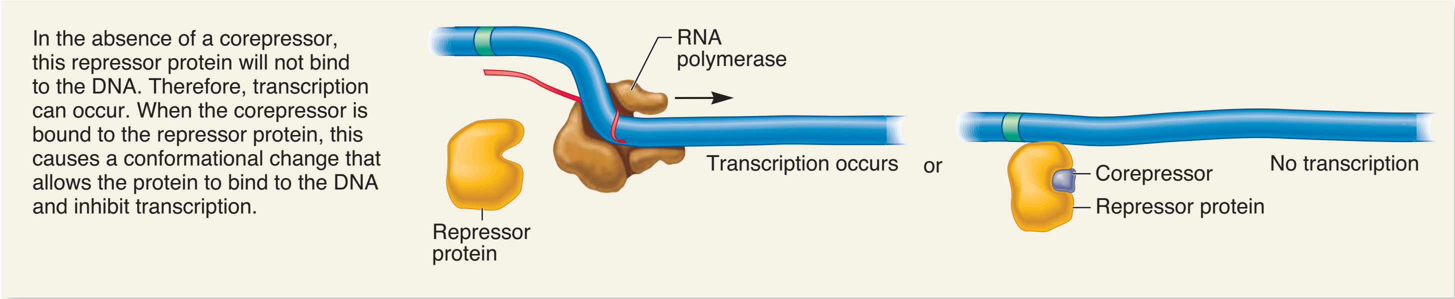
(a) Repressor protein, inducer molecule, inducible gene



This activator protein cannot bind to the DNA unless an inducer is present. When the inducer is bound to the activator protein, this enables the activator protein to bind to the DNA and activate transcription.

No transcription
Activator protein
or
Inducer
RNA polymerase
Transcription occurs
Activator protein

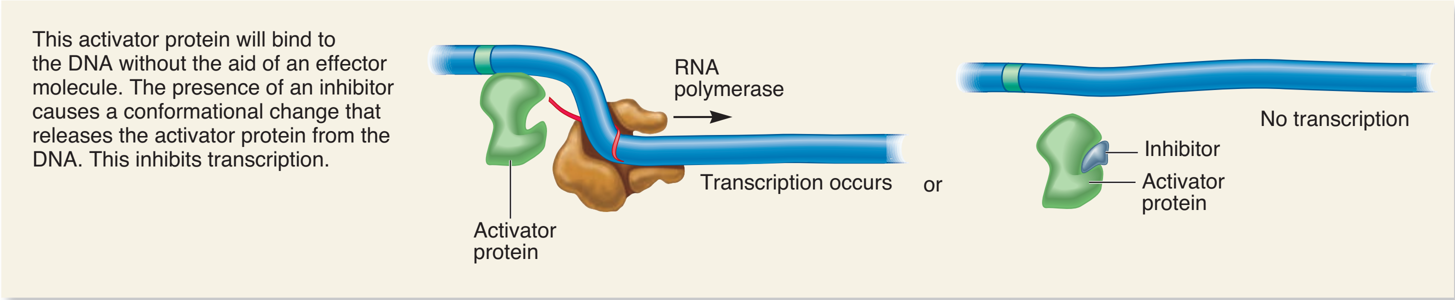
(b) Activator protein, inducer molecule, inducible gene



In the absence of a corepressor, this repressor protein will not bind to the DNA. Therefore, transcription can occur. When the corepressor is bound to the repressor protein, this causes a conformational change that allows the protein to bind to the DNA and inhibit transcription.

RNA polymerase
Transcription occurs
Repressor protein
or
Corepressor
No transcription
Repressor protein

(c) Repressor protein, corepressor molecule, repressible gene



This activator protein will bind to the DNA without the aid of an effector molecule. The presence of an inhibitor causes a conformational change that releases the activator protein from the DNA. This inhibits transcription.

RNA polymerase
Transcription occurs
Activator protein
or
No transcription
Inhibitor
Activator protein

(d) Activator protein, inhibitor molecule, repressible gene

FIGURE 16.2 **Binding sites on a genetic regulatory protein.** In these examples, a regulatory protein has two binding sites: one for a small effector molecule and one for DNA. The binding of the small effector molecule changes the conformation of the regulatory protein, which alters the DNA-binding site structure, thereby influencing whether the protein can bind to the DNA.

Concept check: *Which of these are genetic regulatory proteins and which are small effector molecules?*

16.2 REGULATION OF THE *lac* OPERON

We will now turn our attention to a specific example of gene regulation that is found in *E. coli*. This example involves genes that play a role in the utilization of lactose, which is a sugar found in milk. As you will learn, the regulation of these genes involves a repressor protein and also an activator protein.

The Phenomenon of Enzyme Adaptation Is Due to the Synthesis of Cellular Proteins

Our initial understanding of gene regulation can be traced back to the creative minds of François Jacob and Jacques Monod at the Pasteur Institute in Paris, France. Their research into genes and gene regulation stemmed from an interest in the phenomenon known as **enzyme adaptation,** which had been identified at the turn of the twentieth century. Enzyme adaptation refers to the observation that a particular enzyme appears within a living cell only after the cell has been exposed to the substrate for that enzyme. When a bacterium is not exposed to a particular substance, it does not make the enzymes needed to metabolize that substance.

To investigate this phenomenon, Jacob and Monod focused their attention on lactose metabolism in *E. coli*. Key experimental observations that led to an understanding of this genetic system are listed here.

1. The exposure of bacterial cells to lactose increased the levels of lactose-utilizing enzymes by 1000- to 10,000-fold.
2. Antibody and labeling techniques revealed that the increase in the activity of these enzymes was due to the increased synthesis of the enzymes.
3. The removal of lactose from the environment caused an abrupt termination in the synthesis of the enzymes.
4. Mutations that prevented the synthesis of particular proteins involved in lactose utilization showed that a separate gene encoded each protein.

These critical observations indicated to Jacob and Monod that enzyme adaptation is due to the synthesis of specific cellular proteins in response to lactose in the environment. Next, we will examine how Jacob and Monod discovered that this phenomenon is controlled by the interactions between genetic regulatory proteins and small effector molecules. In other words, we will see that enzyme adaptation is due to the transcriptional regulation of genes.

The *lac* Operon Encodes Proteins Involved in Lactose Metabolism

In bacteria, it is common for a few genes to be arranged together in an **operon**—a group of two or more genes under the transcriptional control of a single promoter. An operon encodes a **polycistronic RNA,** an RNA that contains the sequences of two or more genes. Why do operons occur in bacteria? One biological advantage of an operon organization is that it allows a bacterium to coordinately regulate a group of genes that are involved with a common functional goal; the expression of the genes occurs as a single unit. For transcription to take place, an operon is flanked by a **promoter** that signals the beginning of transcription and a **terminator** that specifies the end of transcription. Two or more genes are found between these two sequences.

Figure 16.3a shows the organization of the genes involved in lactose utilization and their transcriptional regulation. Two distinct transcriptional units are present. The first unit, known as the *lac* operon, contains a CAP site; promoter (*lacP*); operator site (*lacO*); three structural genes, *lacZ, lacY,* and *lacA;* and a terminator. *LacZ* encodes the enzyme β-galactosidase, an enzyme that cleaves lactose into galactose and glucose. As a side reaction, β-galactosidase also converts a small percentage of lactose into allolactose, a structurally similar sugar (**Figure 16.3b**). As we will see later, allolactose acts as a small effector molecule to regulate the *lac* operon. The *lacY* gene encodes lactose permease, a membrane protein required for the active transport of lactose into the cytoplasm of the bacterium. The *lacA* gene encodes galactoside transacetylase, an enzyme that covalently modifies lactose and lactose analogs. Although the functional necessity of the transacetylase remains unclear, the acetylation of nonmetabolizable lactose analogs may prevent their toxic buildup within the bacterial cytoplasm.

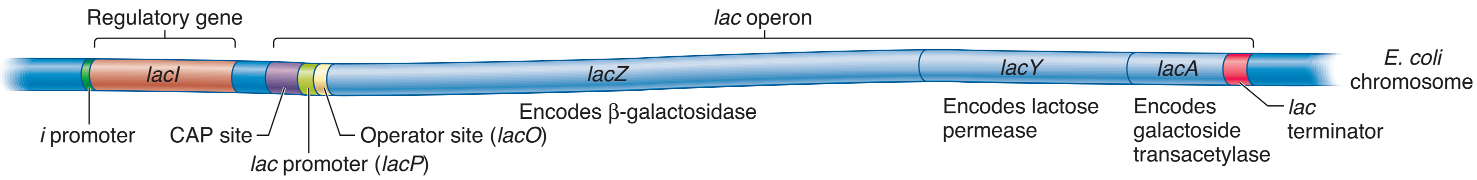
The CAP site and the operator site are short DNA segments that function in gene regulation. The **CAP site** is a DNA sequence recognized by an activator protein called the **catabolite activator protein (CAP).** The **operator site** (also known simply as the **operator**) is a sequence of bases that provides a binding site for a repressor protein.

A second transcriptional unit involved in genetic regulation is the *lacI* gene (see Figure 16.3a), which is not part of the *lac* operon. The *lacI* gene, which is constitutively expressed at fairly low levels, has its own promoter, the *i* promoter. The *lacI* gene encodes the **lac repressor,** a protein that is important for the regulation of the *lac* operon. The lac repressor functions as a homotetramer, a protein composed of four identical subunits. The amount of lac repressor made is approximately 10 proteins per cell. Only a small amount of the lac repressor protein is needed to repress the *lac* operon.

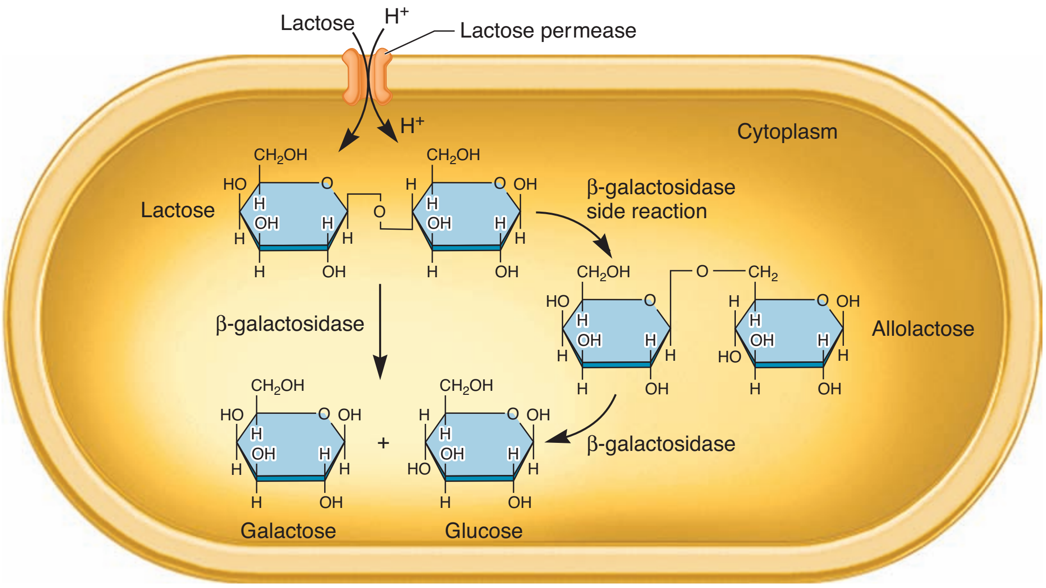
The *lac* Operon Is Regulated by a Repressor Protein

The *lac* operon can be transcriptionally regulated in more than one way. The first mechanism we will examine is one that is inducible and under negative control. As shown in **Figure 16.4**, this form of regulation involves the lac repressor protein, which binds to the sequence of nucleotides found within the *lac* operator site. Once bound, the lac repressor prevents RNA polymerase from transcribing the *lacZ*, *lacY*, and *lacA* genes (Figure 16.4a). The binding of the repressor to the operator site is a reversible process. In the absence of allolactose, the lac repressor is bound to the operator site most of the time.

The ability of the lac repressor to bind to the operator site depends on whether or not allolactose is bound to it. Each of the repressor protein's four subunits has a single binding site for allolactose, the inducer. How does a small molecule like allolactose exert its effects? When allolactose binds to the repressor, a


Regulatory gene
lac operon
lacI
lacZ
lacY
lacA
E. coli chromosome
i promoter
CAP site
lac promoter (*lacP*)
Operator site (*lacO*)
Encodes β-galactosidase
Encodes lactose permease
Encodes galactoside transacetylase
lac terminator

(a) Organization of DNA sequences in the *lac* region of the *E. coli* chromosome


Lactose
H+
Lactose permease
H+
Cytoplasm
Lactose
β-galactosidase side reaction
Allolactose
β-galactosidase
β-galactosidase
Galactose
Glucose

(b) Functions of lactose permease and β-galactosidase

ONLINE ANIMATION

FIGURE 16.3 **Organization of the *lac* operon and other genes involved with lactose metabolism in *E. coli.*** **(a)** The CAP site is the binding site for the catabolite activator protein (CAP). The operator site is a binding site for the lac repressor. The promoter (*lacP*) is responsible for the transcription of the *lacZ*, *lacY,* and *lacA* genes as a single unit, which ends at the *lac* terminator. The *i* promoter is responsible for the transcription of the *lacI* gene. **(b)** Lactose permease allows the uptake of lactose into the bacterial cytoplasm. It cotransports lactose with H^+. Because bacteria maintain an H^+ gradient across their cytoplasmic membrane, this cotransport permits the active accumulation of lactose against a gradient. β-galactosidase is a cytoplasmic enzyme that cleaves lactose and related compounds into galactose and glucose. As a minor side reaction, β-galactosidase also converts lactose into allolactose. Allolactose can also be broken down into galactose and glucose.

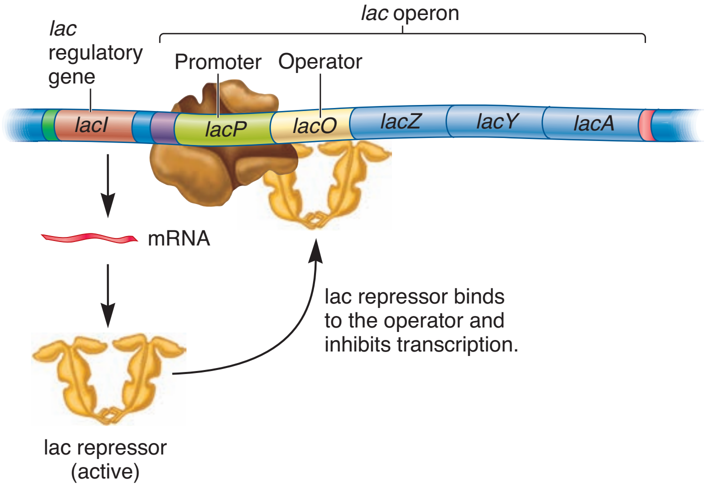
Concept check: *Which genes are under the control of the lac promoter?*

conformational change occurs in the lac repressor protein that prevents it from binding to the operator site. Under these conditions, RNA polymerase is now free to transcribe the operon (Figure 16.4b). In genetic terms, we would say that the operon has been **induced.** The action of a small effector molecule, such as allolactose, is called **allosteric regulation.** Allosteric proteins have at least two binding sites. The effector molecule binds to the protein's **allosteric site,** which is a site other than the protein's active site. In the case of the lac repressor, the active site is the part of the protein that binds to the DNA.

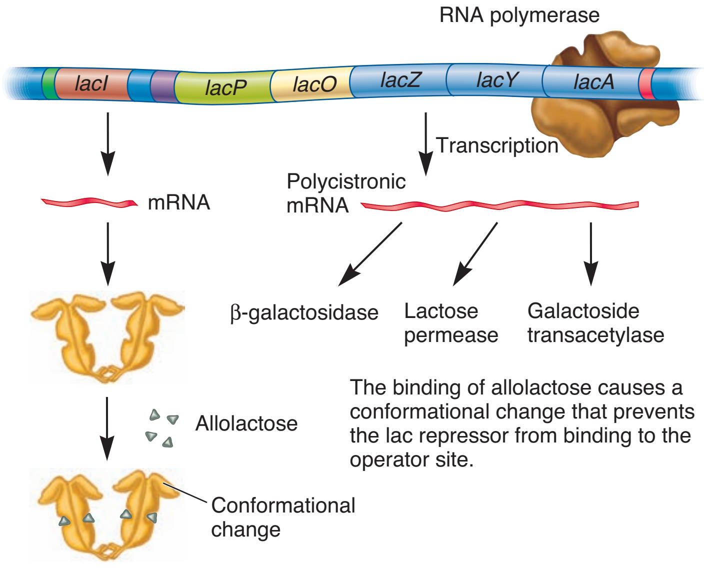
Rare mutations in the *lacI* gene that alter the regulation of the *lac* operon reveal that the lac repressor is composed of a protein domain that binds to the DNA and another domain that contains the allolactose-binding site. As discussed later in Figure 16.7, researchers have identified $lacI^-$ mutations that result in the constitutive expression of the *lac* operon, which means that it is expressed in the presence and absence of lactose. Such mutations may result in an inability to synthesize any repressor protein or they may produce a repressor protein that is unable to bind to the DNA at the *lac* operator site. If the lac repressor is unable to bind to the DNA, the *lac* operon cannot be repressed. By comparison, $lacI^S$ mutations have the opposite effect—the *lac* operon cannot be induced even in the presence of lactose. These mutations, which are called super-repressor mutations, are typically located in the domain that binds allolactose. The mutation usually results in a lac repressor protein that cannot bind allolactose. If the lac repressor is unable to bind allolactose, it will remain bound to the *lac* operator site and therefore induction cannot occur.

The Regulation of the *lac* Operon Allows a Bacterium to Respond to Environmental Change

To better appreciate *lac* operon regulation at the cellular level, let's consider the process as it occurs over time. **Figure 16.5**


lac operon
lac regulatory gene
Promoter
Operator
lacI
lacP
lacO
lacZ
lacY
lacA
mRNA
lac repressor binds to the operator and inhibits transcription.
lac repressor (active)

(a) No lactose in the environment


RNA polymerase
lacI
lacP
lacO
lacZ
lacY
lacA
Transcription
mRNA
Polycistronic mRNA
β-galactosidase
Lactose permease
Galactoside transacetylase
The binding of allolactose causes a conformational change that prevents the lac repressor from binding to the operator site.
Allolactose
Conformational change

(b) Lactose present

ONLINE ANIMATION

FIGURE 16.4 **Mechanism of induction of the *lac* operon.** **(a)** In the absence of the inducer allolactose, the repressor protein is tightly bound to the operator site, thereby inhibiting the ability of RNA polymerase to transcribe the operon. **(b)** When allolactose is available, it binds to the repressor. This alters the conformation of the repressor protein in a way that prevents it from binding to the operator site. Therefore, RNA polymerase can transcribe the operon. (The CAP site is not labeled in this drawing.)

illustrates the effects of external lactose on the regulation of the *lac* operon. In the absence of lactose, no inducer is available to bind to the lac repressor. Therefore, the lac repressor binds to the operator site and inhibits transcription. In reality, the repressor does not completely inhibit transcription, so very small amounts of β-galactosidase, lactose permease, and transacetylase are made. However, the levels are far too low for the bacterium to readily use lactose. When the bacterium is exposed to lactose, a small amount can be transported into the cytoplasm via lactose permease, and β-galactosidase converts some of it to allolactose. As this occurs, the cytoplasmic level of allolactose gradually rises; eventually, allolactose binds to the lac repressor. The binding of allolactose promotes a conformational change that prevents the repressor from binding to the *lac* operator site, thereby allowing transcription of the *lacZ, lacY,* and *lacA* genes to occur. Translation of the encoded polypeptides produces the proteins needed for lactose uptake and metabolism.

To understand how the induction process is shut off in a lactose-depleted environment, let's consider the interaction between allolactose and the lac repressor. The lac repressor has a measurable affinity for allolactose. The binding of allolactose to the lac repressor is reversible. The likelihood that allolactose will bind to the repressor depends on the allolactose concentration. During induction of the operon, the concentration of allolactose rises and approaches the affinity for the repressor protein. This makes it likely that allolactose will bind to the lac repressor, thereby causing it to be released from the operator site. Later on, however, the bacterial cell metabolizes the sugars, thereby lowering the concentration of allolactose below its affinity for the repressor. At this point, the lac repressor is unlikely to be bound to allolactose. When allolactose is released, the lac repressor returns to the conformation that binds to the operator site. In this way, the binding of the repressor shuts down the *lac* operon when lactose is depleted from the environment. The mRNA and proteins encoded by the *lac* operon are eventually degraded (see Figure 16.5).

The *lacI* Gene Encodes a Diffusible Repressor Protein

Now that we understand the function of the *lac* operon, let's consider one of the experimental approaches that was used to elucidate its regulation. In the 1950s, Jacob, Monod, and their colleague Arthur Pardee had identified a few rare mutant strains of bacteria that had abnormal lactose adaptation. As mentioned earlier, one type of mutant, designated *lacI*$^-$, resulted in the constitutive expression of the *lac* operon even in the absence of lactose. As shown in **Figure 16.6a**, the correct explanation is that a loss-of-function mutation in the *lacI* gene prevented the lac repressor protein from binding to the *lac* operator site and inhibiting transcription. At the time of their work, however, the function of the lac repressor was not yet known. Instead, the researchers incorrectly hypothesized that the *lacI*$^-$ mutation resulted in the synthesis of an internal inducer, making it unnecessary for cells to be exposed to lactose for induction (**Figure 16.6b**).

To further explore the nature of this mutation, Jacob, Monod, and Pardee applied a genetic approach. In order to understand their approach, let's briefly consider the process of bacterial conjugation (described in Chapter 9). The earliest studies of Jacob, Monod, and Pardee in 1959 involved matings between recipient cells, termed F$^-$, and donor cells, which were *Hfr* strains that transferred a portion of the bacterial

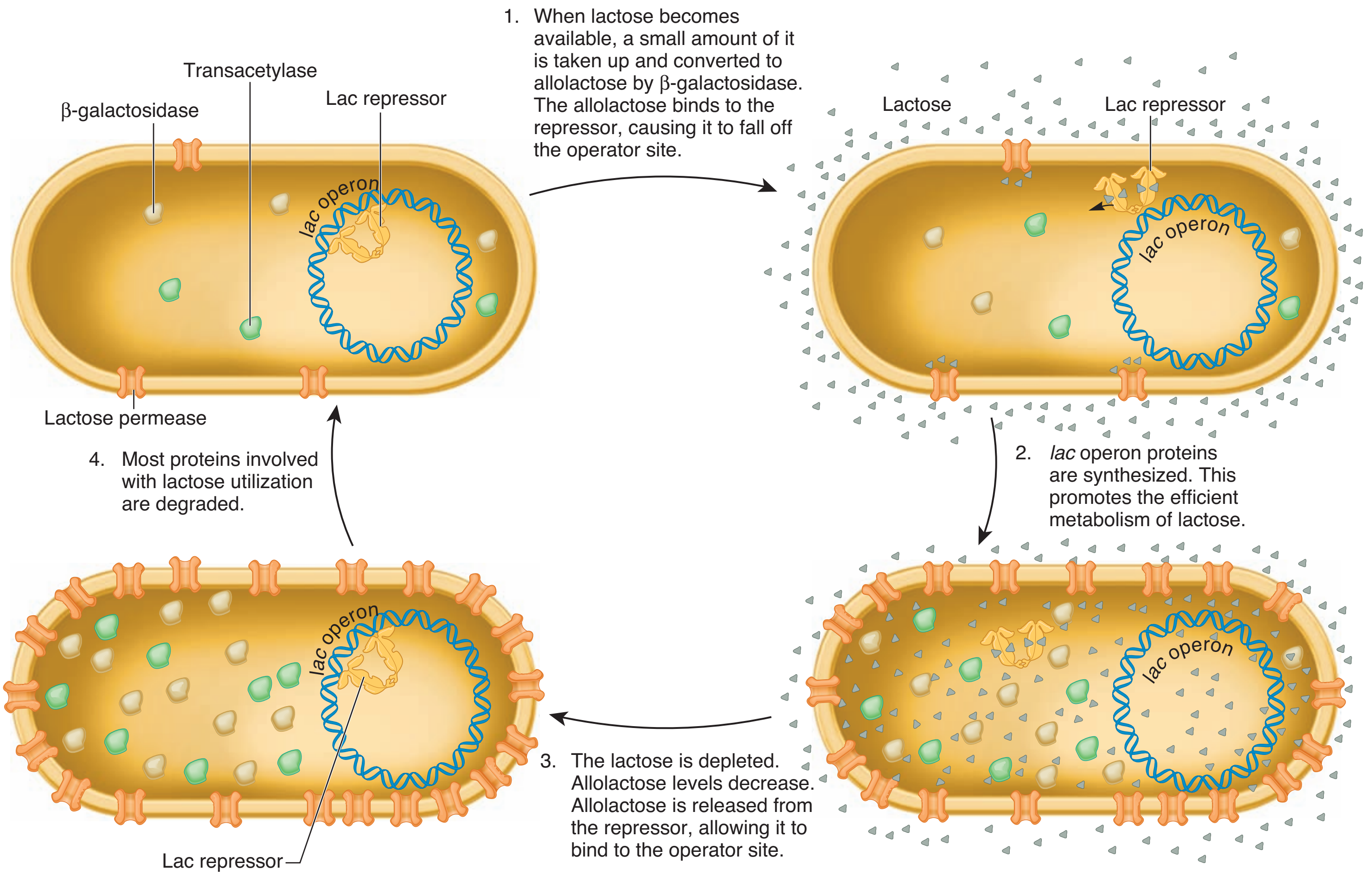
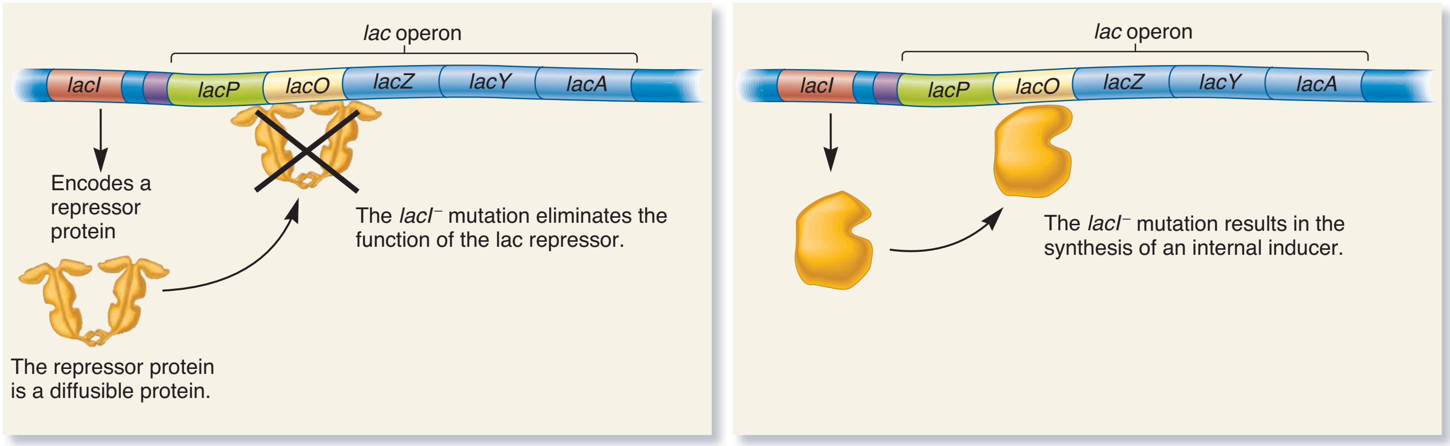

Transacetylase
β-galactosidase
Lac repressor
lac operon
Lactose permease
1. When lactose becomes available, a small amount of it is taken up and converted to allolactose by β-galactosidase. The allolactose binds to the repressor, causing it to fall off the operator site.
Lactose
Lac repressor
lac operon
2. *lac* operon proteins are synthesized. This promotes the efficient metabolism of lactose.
lac operon
3. The lactose is depleted. Allolactose levels decrease. Allolactose is released from the repressor, allowing it to bind to the operator site.
lac operon
Lac repressor
4. Most proteins involved with lactose utilization are degraded.

FIGURE 16.5 **The cycle of *lac* operon induction and repression.**

Genes→Traits The genes and the genetic regulation of the *lac* operon provide the bacterium with the trait of being able to metabolize lactose in the environment. When lactose is present, the genes of the *lac* operon are induced, and the bacterial cell can efficiently metabolize this sugar. When lactose is absent, these genes are repressed, so the bacterium does not waste its energy expressing them. Note: The proteins involved with lactose utilization are fairly stable but they will eventually be degraded.

Concept check: *Under what conditions is the lac repressor bound to the lac operon?*


lac operon
lacI
lacP
lacO
lacZ
lacY
lacA
Encodes a repressor protein
The *lacI*⁻ mutation eliminates the function of the lac repressor.
The repressor protein is a diffusible protein.
lac operon
lacI
lacP
lacO
lacZ
lacY
lacA
The *lacI*⁻ mutation results in the synthesis of an internal inducer.

(a) Correct explanation **(b) Internal inducer hypothesis**

FIGURE 16.6 **Alternative hypotheses to explain how a *lacI*⁻ mutation could cause the constitutive expression of the *lac* operon.** **(a)** The correct explanation in which the *lacI*⁻ mutation eliminates the function of the lac repressor protein, which prevents it from repressing the *lac* operon. **(b)** The internal inducer hypothesis of Jacob, Monod, and Pardee. In this case, the *lacI*⁻ mutation results in the synthesis of an internal inducer that turns on the *lac* operon.

chromosome. Later experiments in 1961 involved the transfer of circular segments of DNA known as F factors. We consider the latter type of experiment here. Sometimes an F factor also carries genes that were originally found within the bacterial chromosome. These types of F factors are called F′ factors (F prime factors). In their studies, Jacob, Monod, and Pardee identified F′ factors that carried the *lacI* gene and the *lac* operon. These F′ factors can be transferred from one cell to another by bacterial conjugation. A strain of bacteria containing F′ factor genes is called a **merozygote,** or partial diploid.

The production of merozygotes was instrumental in allowing Jacob, Monod, and Pardee to elucidate the function of the *lacI* gene. This experimental approach has two key points. First, the two *lacI* genes in a merozygote may be different alleles. For example, the *lacI* gene on the chromosome may be a *lacI*$^-$ allele that causes constitutive expression, whereas the *lacI* gene on the F′ factor may be normal. Second, the genes on the F′ factor and the genes on the bacterial chromosome are not physically adjacent to each other. As we now know, the expression of the *lacI* gene on an F′ factor should produce repressor proteins that can diffuse within the cell and eventually bind to the operator site of the *lac* operon located on the chromosome and also to the operator site on an F′ factor.

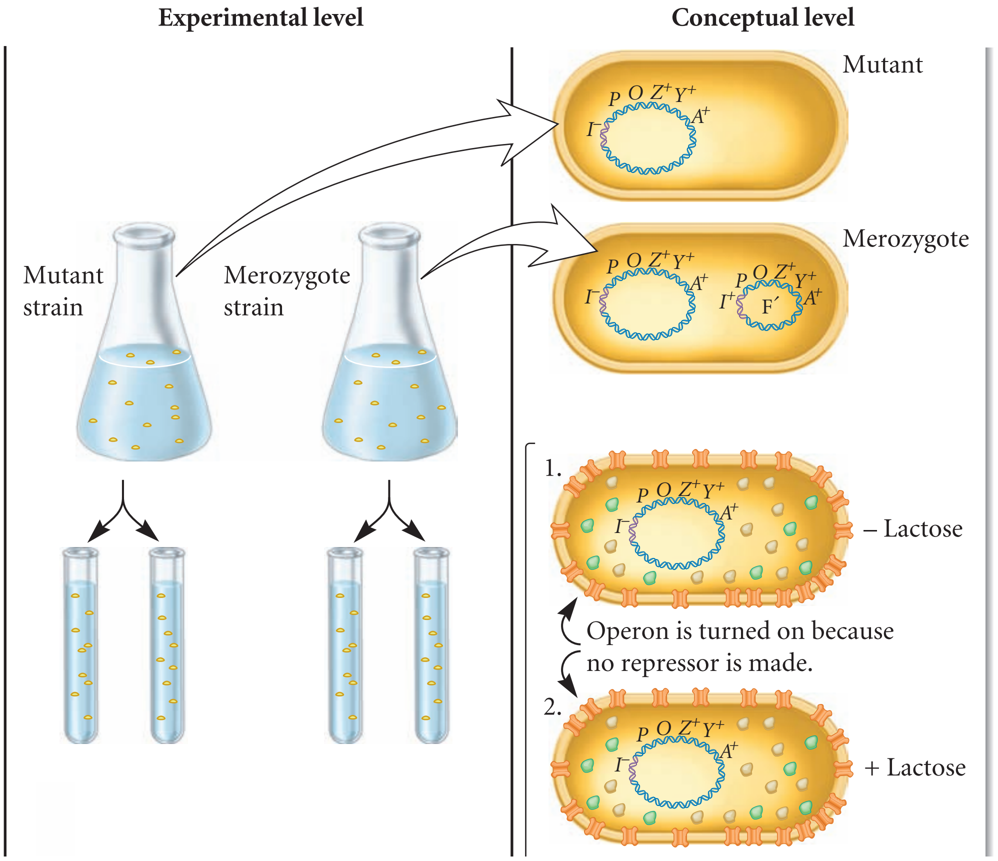
Figure 16.7 shows one experiment of Jacob, Monod, and Pardee in which they analyzed a *lacI*$^-$ mutant strain that was already known to constitutively express the *lac* operon and compared it to the corresponding merozygote. The merozygote had a *lacI*$^-$ mutant gene on the chromosome and a normal *lacI* gene on an F′ factor. These two strains were grown and then divided into two tubes each. In half of the tubes, lactose was omitted. In the other tubes, the strains were incubated with lactose to determine if lactose was needed to induce the expression of the operon. The cells were lysed by sonication, and then a lactose analog, β-*o*-nitrophenylgalactoside (β-ONPG), was added. This molecule is colorless, but β-galactosidase cleaves it into a product that has a yellow color. Therefore, the amount of yellow color produced in a given amount of time is a measure of the amount of β-galactosidase that is being expressed from the *lac* operon.

▶ THE HYPOTHESIS

The *lacI*$^-$ mutation results in the synthesis of an internal inducer.

▶ TESTING THE HYPOTHESIS — FIGURE 16.7 **Evidence that the *lacI* gene encodes a diffusible repressor protein.**

Starting material: The genotype of the mutant strain was *lacI*$^-$ *lacZ*$^+$ *lacY*$^+$ *lacA*$^+$. The merozygote strain had an F′ factor that was *lacI*$^+$ *lacZ*$^+$ *lacY*$^+$ *lacA*$^+$, which had been introduced into the mutant strain via conjugation.



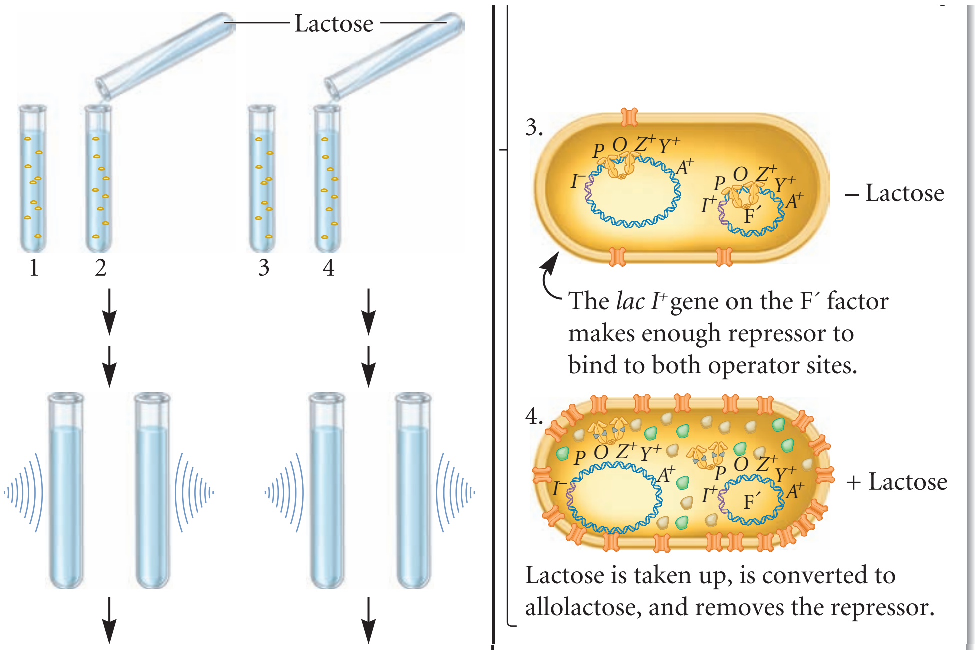
1. Grow mutant strain and merozygote strain separately.

2. Divide each strain into two tubes.

3. In one of the two tubes, add lactose.

4. Incubate the cells long enough to allow *lac* operon induction.

5. Burst the cells with a sonicator. This allows β-galactosidase to escape from the cells.


Lactose
1
2
3
4
3.
P O Z^+ Y^+ A^+
I^-
P O Z^+ Y^+ A^+
I^+
F′
– Lactose
4.
P O Z^+ Y^+ A^+
I^-
P O Z^+ Y^+ A^+
I^+
F′
+ Lactose

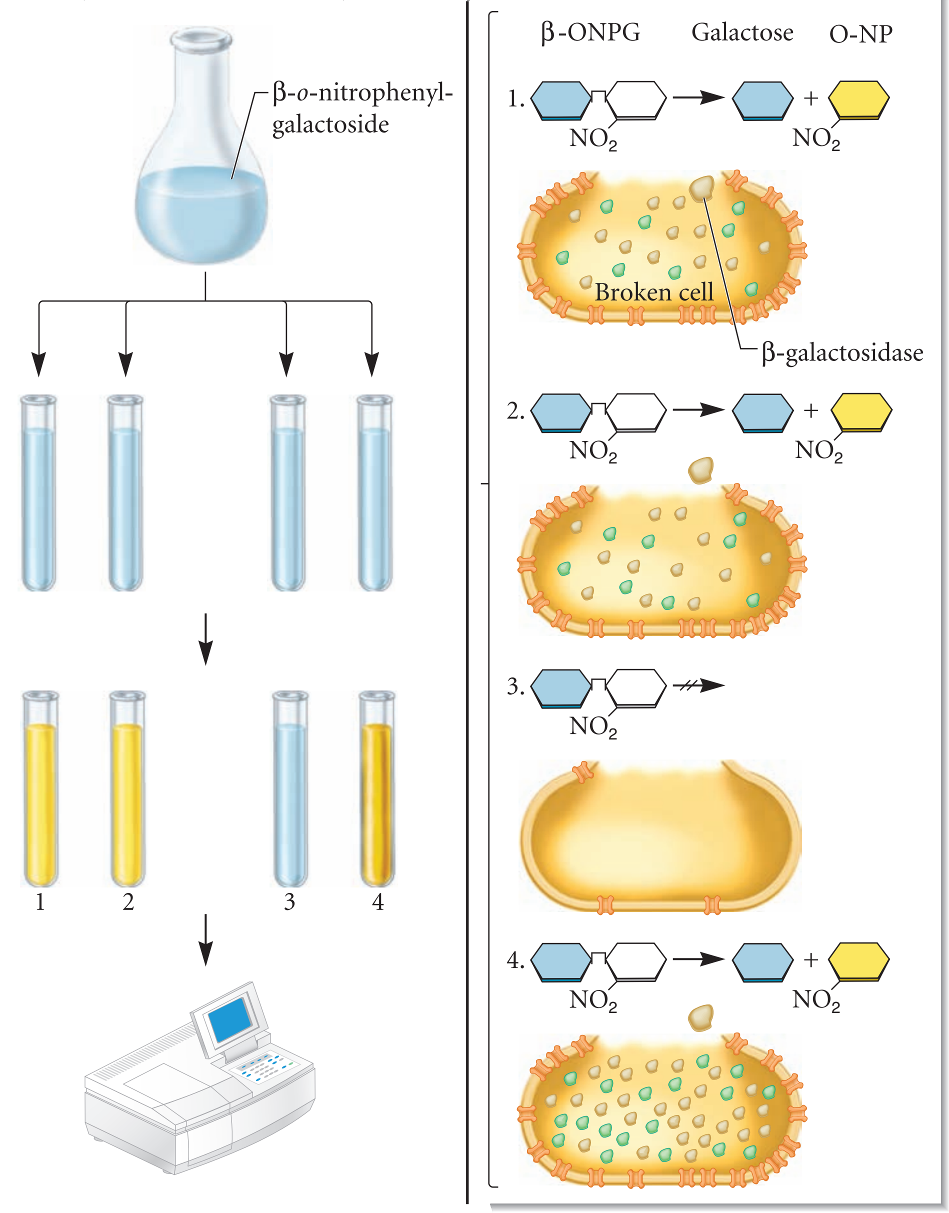
The *lac* I^+ gene on the F′ factor makes enough repressor to bind to both operator sites.

Lactose is taken up, is converted to allolactose, and removes the repressor.

6. Add β-*o*-nitrophenylgalactoside (β-ONPG). This is a colorless compound. β-galactosidase will cleave the compound to produce galactose and *o*-nitrophenol (O-NP). O-NP has a yellow color. The deeper the yellow color, the more β-galactosidase was produced.

7. Incubate the sonicated cells to allow β-galactosidase time to cleave β-ONPG.

8. Measure the yellow color produced with a spectrophotometer. (See the Appendix for a description of spectrophotometry.)


β-*o*-nitrophenyl-galactoside
β-ONPG
Galactose
O-NP
1.
NO_2
+
NO_2
Broken cell
β-galactosidase
2.
NO_2
+
NO_2
3.
NO_2
4.
NO_2
+
NO_2
1
2
3
4

▶ THE DATA

Strain	*Addition of Lactose*	*Amount of β-Galactosidase (percentage of parent strain)*
Mutant	No	100%
Mutant	Yes	100%
Merozygote	No	<1%
Merozygote	Yes	220%

Data from F. Jacob, and J. Monod (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol. 3,* 318–356.

▶ INTERPRETING THE DATA

As seen in the data, the yellow color production in the original mutant strain was the same in the presence or absence of lactose. This result is expected because the expression of β-galactosidase in the *lacI*$^-$ mutant strain was already known to be constitutive. In other words, the presence of lactose was not needed to induce the operon due to a defective *lacI* gene. In the merozygote strain, however, a different result was obtained. In the absence of lactose, the *lac* operons were repressed—even the operon on the bacterial chromosome. How do we explain these results? Because the normal *lacI* gene on the F′ factor was not physically located next to the chromosomal *lac* operon, this result is consistent with the idea that the *lacI* gene codes for a repressor protein that can diffuse throughout the cell and bind to any *lac* operon. The hypothesis that the *lacI*$^-$ mutation resulted in the synthesis of an internal inducer was rejected. If that hypothesis had been correct, the merozygote strain would have still made an internal inducer, and the *lac* operons in the merozygote would have been expressed in the absence of lactose. This result was not obtained.

The interactions between regulatory proteins and DNA sequences illustrated in this experiment have led to the definition of two genetic terms. A ***trans*-effect** is a form of genetic regulation that can occur even though two DNA segments are not physically adjacent. The action of the lac repressor on the *lac* operon is a *trans*-effect. A regulatory protein, such as the lac repressor, is called a ***trans*-acting factor.** In contrast, a ***cis*-acting element** is a DNA segment that must be adjacent to the gene(s) that it regulates, and it is said to have a ***cis*-effect** on gene expression. The *lac* operator site is an example of a *cis*-acting element. A *trans*-effect is mediated by genes that encode regulatory proteins, whereas a *cis*-effect is mediated by DNA sequences that are bound by regulatory proteins.

Jacob and Monod also isolated constitutive mutants that affected the operator site, *lacO*. **Table 16.1** summarizes the effects of mutations based on their locations in the *lacI* regulatory gene versus *lacO* and their analysis in merozygotes. As seen here, a loss-of-function mutation in a gene encoding a repressor protein has the same effect as a mutation in an operator site that cannot bind a repressor protein. In both cases, the genes of the *lac* operon are constitutively expressed. In a merozygote, however, the results are quite different. When a normal *lacI* gene and a normal *lac* operon are introduced into a cell harboring a defective *lacI* gene, the normal *lacI* gene can regulate both operons. In contrast, when a *lac* operon with a normal operator site is introduced into a cell with a defective operator site, the operon with the defective operator site continues to be expressed without lactose present. Overall, a mutation in a *trans*-acting factor can be complemented by the introduction of a second gene with normal function. However, a mutation in a *cis*-acting element is not affected by the introduction of another *cis*-acting element with normal function into the cell.

TABLE **16.1**

A Comparison of Loss-of-Function Mutations in the *lacI* Gene Versus the Operator Site

		Expression of the *lac* Operon (%)	
Chromosome	**F′ factor**	**With Lactose**	**Without Lactose**
Wild type	None	100	<1
lacI$^-$	None	100	100
lacO$^-$	None	100	100
lacI$^-$	*lacI*$^+$ and a normal *lac* operon	200	<1
lacO$^-$	*lacI*$^+$ and a normal *lac* operon	200	100

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

The *lac* Operon Is Also Regulated by an Activator Protein

The *lac* operon can be transcriptionally regulated in a second way, known as **catabolite repression.** (As we will see, this is a somewhat imprecise term.) This form of transcriptional regulation is influenced by the presence of glucose, which is a catabolite—a substance that is broken down inside the cell. The presence of glucose ultimately leads to repression of the *lac* operon. When exposed to both glucose and lactose, *E. coli* cells first use glucose, and catabolite repression prevents the use of lactose. Why is this an advantage? The explanation is efficiency. The bacterium does not have to express all of the genes necessary for both glucose and lactose metabolism. If the glucose is used up, catabolite repression is alleviated, and the bacterium then expresses the *lac* operon. The sequential use of two sugars by a bacterium, which is called **diauxic growth,** is a common phenomenon among many bacterial species. Typically, glucose, a more commonly encountered sugar, is metabolized preferentially, and then a second sugar is metabolized only after glucose has been depleted from the environment.

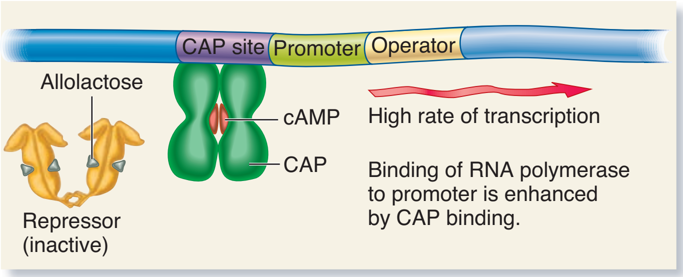
Glucose, however, is not itself the small effector molecule that binds directly to a genetic regulatory protein. Instead, this form of regulation involves a small effector molecule, **cyclic-AMP (cAMP),** which is produced from ATP via an enzyme known as adenylyl cyclase. When a bacterium is exposed to glucose, the transport of glucose into the cell stimulates a signaling pathway that causes the intracellular concentration of cAMP to decrease because the pathway inhibits adenylyl cyclase, the enzyme needed for cAMP synthesis. The effect of cAMP on the *lac* operon is mediated by an activator protein called the catabolite activator protein (CAP). CAP is composed of two subunits, each of which binds one molecule of cAMP.

Figure 16.8 considers how the interplay between the lac repressor and CAP determines whether the *lac* operon is expressed in the presence or absence of lactose and/or glucose. When only lactose is present, cAMP levels are high (Figure 16.8a). The cAMP binds to CAP, and then CAP binds to the CAP site and stimulates the ability of RNA polymerase to begin transcription. In the presence of lactose, the lac repressor is not bound to the operator site, so transcription can proceed at a high rate. In the absence of both lactose and glucose, cAMP levels are also high (Figure 16.8b). Under these conditions, however, the binding of the lac repressor inhibits transcription even though CAP is bound to the DNA. Therefore, the transcription rate is very low.

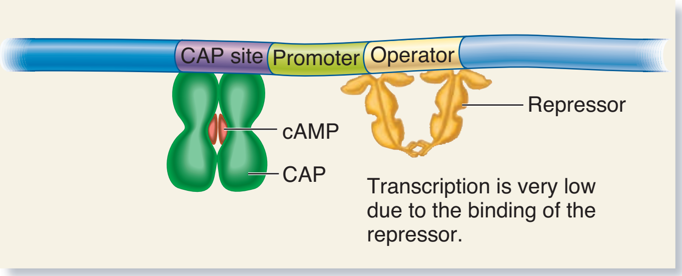
Figure 16.8c considers the situation in which both sugars are present. The presence of lactose causes the lac repressor to be inactive, which prevents it from binding to the operator site. Even so, the presence of glucose decreases cAMP levels so that cAMP is released from CAP, which prevents CAP from binding to the CAP site. Because CAP is not bound to the CAP site, the transcription of the *lac* operon is low in the presence of both sugars. Finally, Figure 16.8d illustrates what happens when only glucose is present. The transcription of the *lac* operon is very low because the lac repressor is bound to the operator site and CAP is not bound to the CAP site due to low cAMP levels.

The effect of glucose, called catabolite repression, may seem like a puzzling way to describe this process because this regulation involves the action of an inducer (cAMP) and an activator protein (CAP), not a repressor. The term was coined before the action of the cAMP-CAP complex was understood. At that time, the primary observation was that glucose (a catabolite) inhibited (repressed) lactose metabolism.

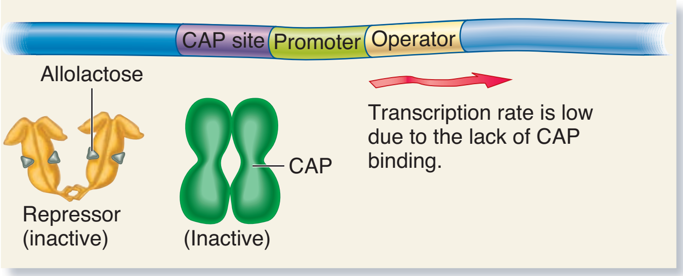
Many bacterial promoters that transcribe genes involved in the breakdown of other sugars, such as maltose, arabinose, and melibiose, also have binding sites for CAP. Therefore, when glucose levels are high, these operons are inhibited. This promotes diauxic growth.


CAP site
Promoter
Operator
Allolactose
cAMP
CAP
Repressor (inactive)
High rate of transcription
Binding of RNA polymerase to promoter is enhanced by CAP binding.

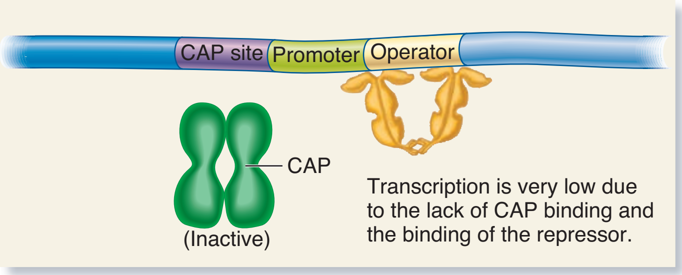
(a) Lactose, no glucose (high cAMP)


CAP site
Promoter
Operator
cAMP
CAP
Repressor
Transcription is very low due to the binding of the repressor.

(b) No lactose or glucose (high cAMP)


CAP site
Promoter
Operator
Allolactose
Repressor (inactive)
CAP
(Inactive)
Transcription rate is low due to the lack of CAP binding.

(c) Lactose and glucose (low cAMP)


CAP site
Promoter
Operator
CAP
(Inactive)
Transcription is very low due to the lack of CAP binding and the binding of the repressor.

(d) Glucose, no lactose (low cAMP)

INTERACTIVE EXERCISE

FIGURE 16.8 The roles of the lac repressor and catabolite activator protein (CAP) in the regulation of the *lac* operon. This figure illustrates how the *lac* operon is regulated depending on its exposure to lactose or glucose.

Genes→Traits The mechanism of catabolite repression provides the bacterium with the trait of being able to choose between two sugars. When exposed to both glucose and lactose, the bacterium chooses glucose first. After the glucose is used up, it then expresses the genes necessary for lactose metabolism. This trait allows the bacterium to more efficiently use sugars from its environment.

Concept check: Why is it beneficial to the bacterium to regulate the lac operon with both a repressor protein and an activator protein?

Further Studies Have Revealed That the *lac* Operon Has Three Operator Sites for the lac Repressor

Our traditional view of the regulation of the *lac* operon has been modified as we gain a greater understanding of the molecular process. In particular, detailed genetic and crystallographic studies have shown that the binding of the lac repressor is more complex than originally realized. The site in the *lac* operon commonly called the operator site was first identified by mutations that prevented lac repressor binding. These mutations, called *lacO*$^-$ or *lacO*^{C} mutants, resulted in the constitutive expression of the *lac* operon even in strains that make a normal lac repressor protein. *LacO*^{C} mutations were localized in the *lac* operator site that is now known as O_1. This led to the view that a single operator site was bound by the lac repressor to inhibit transcription, as in Figure 16.4.

In the late 1970s and 1980s, two additional operator sites were identified. As shown at the top of **Figure 16.9**, these sites are called O_2 and O_3. O_1 is the operator site slightly downstream from the promoter. O_2 is located farther downstream in the *lacZ* coding sequence, and O_3 is located slightly upstream from the promoter. The O_2 and O_3 operator sites were initially called pseudo-operators, because substantial repression occurred in the absence of either one of them. However, studies by Benno Müller-Hill and his colleagues revealed a surprising result. As shown in Figure 16.9 (fourth example down), if both O_2 and O_3 are missing, repression is dramatically reduced even when O_1 is present. When O_1 is missing, even in the presence of one of the other operator sites, repression is nearly abolished.

How were these results interpreted? The data of Figure 16.9 supported a hypothesis that the lac repressor must bind to O_1 and either O_2 or O_3 to cause full repression. According to this view, the lac repressor can readily bind to O_1 and O_2, or to O_1 and O_3, but not to O_2 and O_3. If either O_2 or O_3 were missing, maximal repression is not achieved because it is less likely for the repressor to bind when only two operator sites are present. Look at Figure 16.9 and you will notice that the operator sites are a fair distance from each other. For this reason, it was proposed that the binding of the lac repressor to two operator sites requires the DNA to form a loop. A loop in the DNA would bring the operator sites closer together, thereby facilitating the binding of the repressor protein (**Figure 16.10a**).

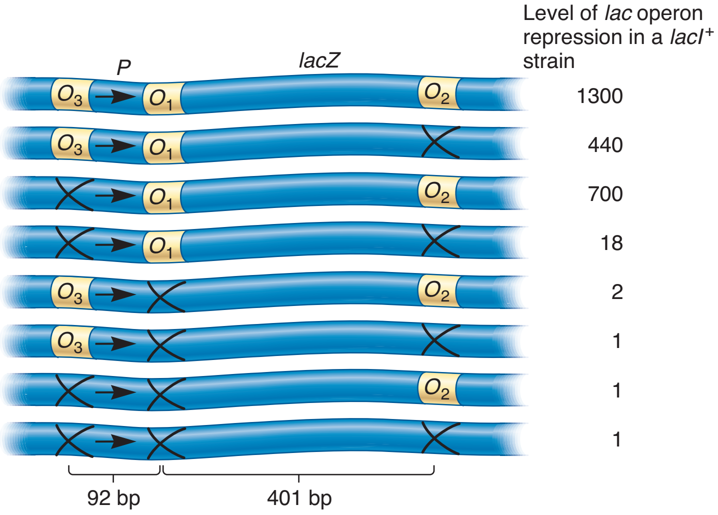


FIGURE 16.9 **The identification of three *lac* operator sites.** The top of this figure shows the locations of three *lac* operator sites, designated O_1, O_2, and O_3. The *lac* operator site shown in previous figures is the same as O_1. The arrows depict the starting site for transcription. Defective operator sites are indicated with an X. When all three operator sites are present, the repression of the *lac* operon is 1300-fold; this means there is 1/1300 the level of expression as compared to when lactose is present. This figure also shows the amount of repression when one or more operator sites are removed. A repression value of 1.0 indicates that no repression is occurring. In other words, a value of 1.0 indicates constitutive expression.

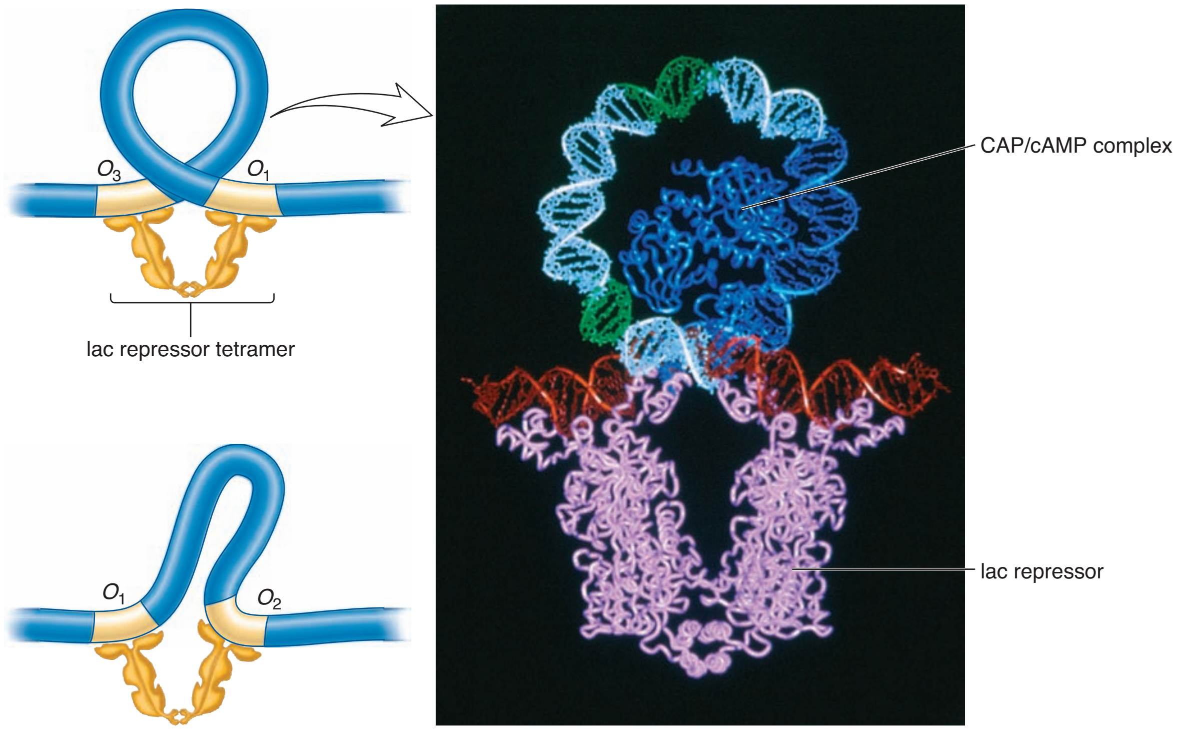
Concept check: *Which data provide the strongest evidence that O_1 is not the only operator site?*

In 1996, the proposal that the lac repressor binds to two operator sites was confirmed by studies in which the lac repressor was crystallized by Mitchell Lewis and his colleagues. The crystal structure of the lac repressor has provided exciting insights into its mechanism of action. As mentioned earlier in this chapter, the lac repressor is a tetramer of four identical subunits. The crystal structure revealed that each dimer within the tetramer recognizes one operator site. **Figure 16.10b** is a molecular model illustrating the binding of the lac repressor to the O_1 and O_3 sites. The amino acid side chains in the protein interact directly with bases in the major groove of the DNA helix. This is how genetic regulatory proteins recognize specific DNA sequences. Because each dimer within the tetramer recognizes a single operator site, the association of two dimers to form a tetramer requires that the two operator sites be close to each other. For this to occur, a loop must form in the DNA. The formation of this loop dramatically inhibits the ability of RNA polymerase to slide past the O_1 site and transcribe the operon.

Figure 16.10b also shows the binding of the cAMP-CAP complex to the CAP site (see the blue protein within the loop). A particularly striking observation is that the binding of the cAMP-CAP complex to the DNA causes a 90° bend in the DNA structure. When the repressor is active—not bound to allolactose—the cAMP-CAP complex facilitates the binding of the lac repressor to the O_1 and O_3 sites. When the repressor is inactive, this bending also appears to be important in the ability of RNA polymerase to initiate transcription slightly downstream from the bend.

16.2 REVIEWING THE KEY CONCEPTS

- Enzyme adaptation is the phenomenon in which the enzymes that metabolize a molecule are made only when the molecule is present in the environment.
- The *lac* operon encodes a polycistronic mRNA for proteins that are involved with the uptake and metabolism of lactose (see Figure 16.3).
- The lac repressor protein binds to the *lac* operator and inhibits transcription. Allolactose binds to the repressor and causes a conformational change that prevents the repressor from binding to the operator. This event induces transcription (see Figure 16.4).
- The regulation of the *lac* operon enables the bacterium *E. coli* to respond to changes in the level of lactose in its environment (see Figure 16.5).
- Jacob and Monod constructed merozygotes to show that the *lacI* gene encodes a diffusible repressor protein (see Figures 16.6, 16.7).
- Mutations in the *lacI* gene and the *lac* operator site may have different effects in a merozygote (see Table 16.1).
- CAP is an activator protein of the *lac* operon. It binds to the CAP site when cAMP levels are high. Glucose inhibits cAMP levels, thereby inhibiting the *lac* operon (see Figure 16.8).



(a) DNA loops caused by the binding of the lac repressor

(b) Proposed model of the lac repressor binding to O_1 and O_3 based on crystallography studies

FIGURE 16.10 **The binding of the lac repressor to two operator sites.** **(a)** The binding of the lac repressor protein to the O_1 and O_3 (top) or to the O_1 and O_2 (bottom) operator sites. Because the two sites are far apart, a loop must form in the DNA. **(b)** A molecular model for the binding of the lac repressor to O_1 and O_3. Each repressor dimer binds to one operator site, so the repressor tetramer brings the two operator sites together. This causes the formation of a DNA loop in the intervening region. Note that the DNA loop contains the –35 and –10 regions (shown in green), which are recognized by σ factor of RNA polymerase (refer back to Figure 14.4). This loop also contains the binding site for the cAMP-CAP complex, which is the blue protein within the loop.

- The lac repressor, which is a tetramer, binds to two operator sites, either to O_1 and O_2, or to O_1 and O_3, causing a loop to form in the DNA (see Figures 16.9, 16.10).

16.2 COMPREHENSION QUESTIONS

1. What is an operon?
   a. A site in the DNA where a regulatory protein binds
   b. A group of genes under the control of a single promoter
   c. An mRNA that encodes several genes
   d. All of the above
2. The binding of __________ to the lac repressor causes the lac repressor to __________ to the operator site and thereby __________ transcription.
   a. glucose, bind, inhibits
   b. allolactose, bind, inhibits
   c. glucose, not bind, increases
   d. allolactose, not bind, increases
3. On its chromosome, an *E. coli* cell is $lacI^- \ lacZ^+ \ lacY^+ \ lacA^+$. It has an F′ factor that is $lacI^+ \ lacZ^+ \ lacY^+ \ lacA^+$. What is the expected level of expression of the *lac* operon genes ($lacZ^+ \ lacY^+ \ lacA^+$) in the absence of lactose?
   a. Both *lac* operons would be expressed.
   b. Neither *lac* operon would be expressed.
   c. Only the chromosomal *lac* operon would be expressed.
   d. Only the *lac* operon on the F′ factor would be expressed.
4. When an *E. coli* cell is exposed to glucose, how does this affect the regulation of the *lac* operon via CAP?
   a. cAMP binds to CAP and transcription is increased.
   b. cAMP binds to CAP and transcription is decreased.
   c. cAMP does not bind to CAP and transcription is increased.
   d. cAMP does not bind to CAP and transcription is decreased.

16.3 REGULATION OF THE *trp* OPERON

We now turn our attention to a second operon in *E. coli* called the *trp* operon, which encodes enzymes involved with the synthesis of the amino acid tryptophan. Like the *lac* operon, the *trp* operon is regulated by a repressor protein. In addition, you will learn that the *trp* operon is regulated by another mechanism called **attenuation** in which transcription begins, but is stopped prematurely, or "attenuated," before most of the *trp* operon is transcribed.

The *trp* Operon Is Regulated by a Repressor Protein

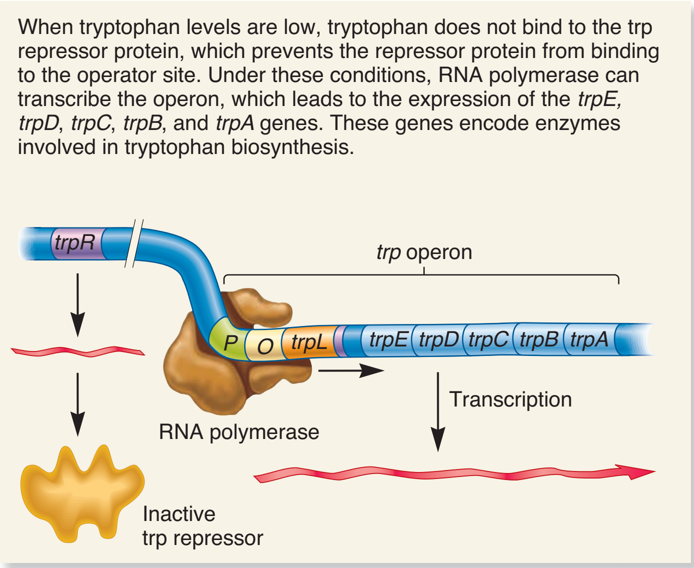
The *trp* operon (pronounced "trip") contains six genes—*trpL, trpE, trpD, trpC, trpB,* and *trpA*. The *trpE, trpD, trpC, trpB,* and *trpA* genes encode enzymes involved in tryptophan biosynthesis, whereas *trpL* plays a regulatory role. The *trpR* gene, which is not part of the *trp* operon, encodes the **trp repressor** protein. When tryptophan levels within the cell are very low, the trp repressor cannot bind to the operator site. Under these conditions, RNA polymerase transcribes the *trp* operon (**Figure 16.11a**). In this way, the cell expresses the genes required for the synthesis of tryptophan. When the tryptophan levels within the cell become high, tryptophan acts as a corepressor that binds to the trp repressor protein. This causes a conformational change in the trp repressor that allows it to bind to the *trp* operator site (**Figure 16.11b**). This inhibits the ability of RNA polymerase to transcribe the operon. Therefore, when a high level of tryptophan is present within the cell—when the cell does not need to make more tryptophan—the *trp* operon is turned off.

The *trp* Operon Is Also Regulated by Attenuation

In the 1970s, after the action of the trp repressor was elucidated, Charles Yanofsky and coworkers made a few unexpected observations. Mutant strains were found that lacked the trp repressor protein. Surprisingly, these mutant strains still could inhibit the expression of the *trp* operon when tryptophan levels were high. In addition, *trp* operon mutations were identified in which a region including the *trpL* gene was missing from the operon. These mutations resulted in higher levels of expression of the other genes in the *trp* operon. As is often the case, unusual observations can lead scientists into productive avenues of study. By pursuing this research further, Yanofsky discovered a second regulatory mechanism in the *trp* operon, called attenuation, that is mediated by the region that includes the *trpL* gene (**Figure 16.11c**).

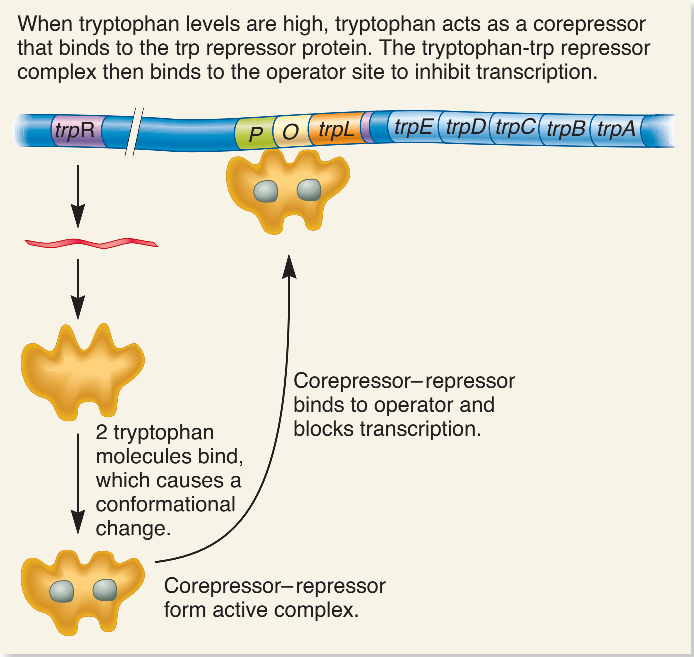
Attenuation can occur in bacteria because the processes of transcription and translation are coupled. As described in Chapter 15, bacterial ribosomes quickly attach to the 5′ end of mRNA soon after its synthesis begins via RNA polymerase. During attenuation, transcription actually begins, but it is terminated before the entire mRNA is made. A segment of DNA, termed the **attenuator sequence,** is important in facilitating this termination. When attenuation occurs, the mRNA from the *trp* operon is made as a short piece that terminates at the attenuator sequence, which is shortly past the *trpL* gene (see Figure 16.11c). Because this short mRNA has been terminated before RNA polymerase has transcribed the *trpE, trpD, trpC, trpB,* and *trpA* genes, it does not encode the proteins required for tryptophan biosynthesis. In this way, attenuation inhibits the further production of tryptophan in the cell.

The segment of the *trp* operon immediately downstream from the operator site plays a critical role during attenuation.



When tryptophan levels are low, tryptophan does not bind to the trp repressor protein, which prevents the repressor protein from binding to the operator site. Under these conditions, RNA polymerase can transcribe the operon, which leads to the expression of the *trpE, trpD, trpC, trpB,* and *trpA* genes. These genes encode enzymes involved in tryptophan biosynthesis.

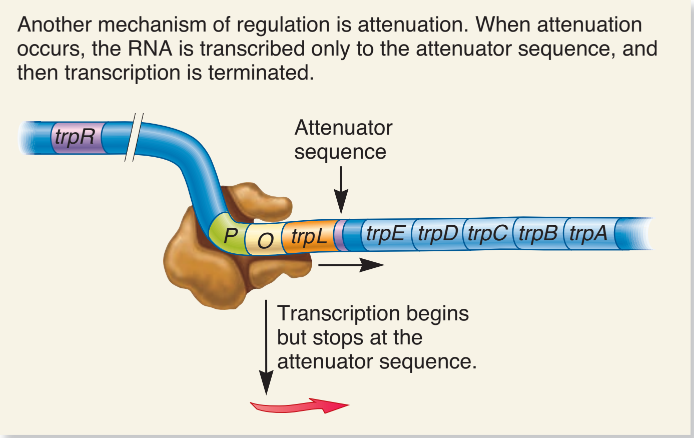
(a) Low tryptophan levels, transcription of the entire *trp* operon occurs



When tryptophan levels are high, tryptophan acts as a corepressor that binds to the trp repressor protein. The tryptophan-trp repressor complex then binds to the operator site to inhibit transcription.

trpR
P
O
trpL
trpE
trpD
trpC
trpB
trpA
Corepressor–repressor binds to operator and blocks transcription.
2 tryptophan molecules bind, which causes a conformational change.
Corepressor–repressor form active complex.

(b) High tryptophan levels, repression occurs



Another mechanism of regulation is attenuation. When attenuation occurs, the RNA is transcribed only to the attenuator sequence, and then transcription is terminated.

(c) High tryptophan levels, attenuation occurs

FIGURE 16.11 Organization of the *trp* operon and regulation via the trp repressor protein and attenuation.

Concept check: How does tryptophan affect the function of the trp repressor?

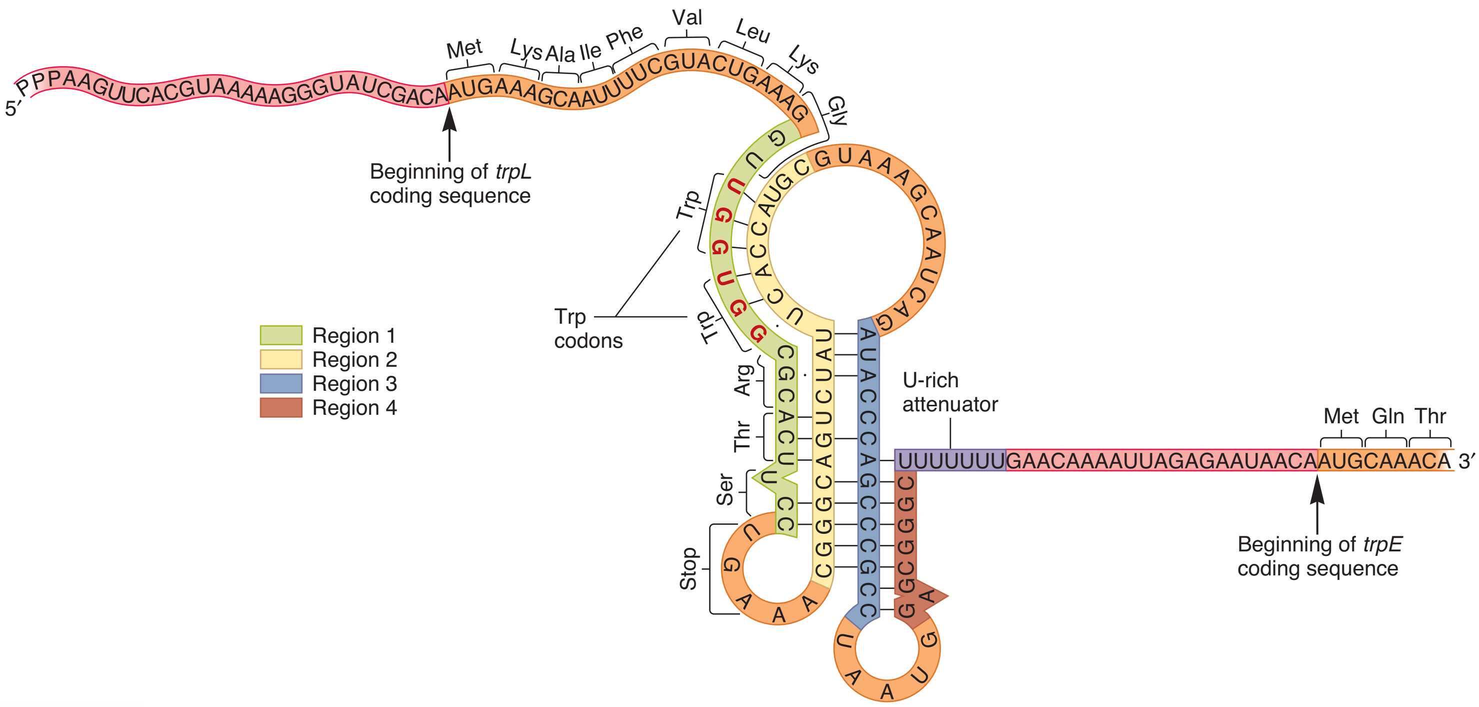

5′ PPPAAGUUCACGUAAAAAGGGUAUCGACAAUGAAAGCAAUUUUCGUACUGAAAG
Met
Lys
Ala
Ile
Phe
Val
Leu
Lys
Gly
Beginning of *trpL* coding sequence
Trp
Trp
Trp codons
Arg
Thr
Ser
Stop
Region 1
Region 2
Region 3
Region 4
U-rich attenuator
UUUUUUUGAACAAAAUUAGAGAAUAACAAUGCAAACA 3′
Met Gln Thr
Beginning of *trpE* coding sequence

FIGURE 16.12 **Sequence of the *trpL* mRNA produced during attenuation.** A second mechanism of regulation is attenuation, which occurs when tryptophan levels are sufficient for protein synthesis. A short mRNA is made that includes the *trpL* coding sequence. As shown here, this mRNA has several regions that are complementary to each other. The possible hydrogen bonding between regions 1 and 2, 2 and 3, and 3 and 4 is also shown. The last U in the purple attenuator sequence is the last nucleotide that is transcribed during attenuation. At low tryptophan concentrations, however, transcription occurs beyond the end of *trpL* and proceeds through the *trpE* gene and the rest of the *trp* operon.

Concept check: *What type of bonding interactions cause the stem-loops to form?*

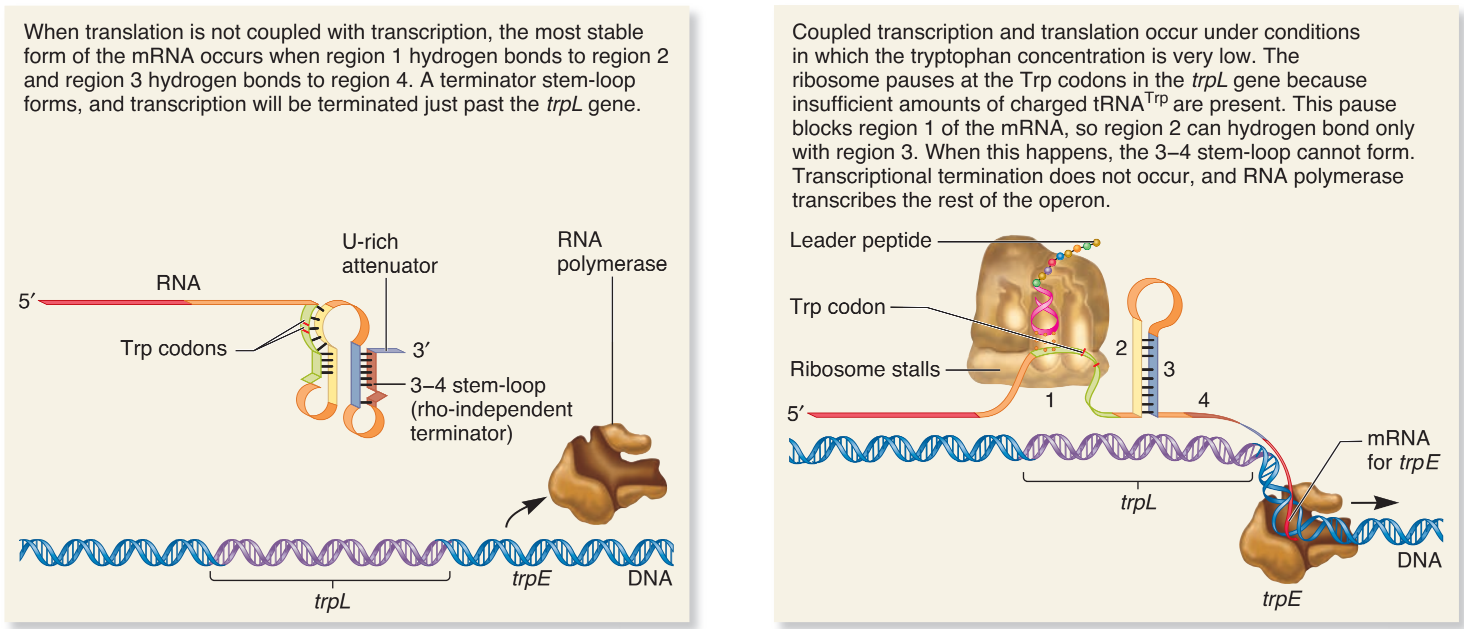
The first gene in the *trp* operon is the *trpL* gene, which encodes a peptide containing 14 amino acids called the leader peptide. As shown in **Figure 16.12**, two features are key in the attenuation mechanism. First, two tryptophan (Trp) codons are found within the mRNA that encodes the *trp* leader peptide. What is the role of these codons? As we will see later, these two codons provide a way to sense whether or not the bacterium has sufficient tryptophan to synthesize its proteins. Second, the mRNA can form stem-loops. The types of stem-loops that form underlie attenuation.

Different combinations of stem-loops are possible due to interactions among four regions within the RNA transcript (see the color key in Figure 16.12). Region 2 is complementary to region 1 and also to region 3. Region 3 is complementary to region 2 as well as to region 4. Therefore, three stem-loops are possible: 1–2, 2–3, and 3–4. Even so, keep in mind that a particular segment of RNA can participate in the formation of only one stem-loop. For example, if region 2 forms a stem-loop with region 1, it cannot (at the same time) form a stem-loop with region 3. Alternatively, if region 2 forms a stem-loop with region 3, then region 3 cannot form a stem-loop with region 4. Though three stem-loops are possible, the 3–4 stem-loop is functionally unique. The 3–4 stem-loop together with the U-rich attenuator sequence acts as an intrinsic terminator—a ρ-independent terminator, as described in Chapter 14. Therefore, the formation of the 3–4 stem-loop causes RNA polymerase to pause, and the U-rich sequence dissociates from the DNA. This terminates transcription at the U-rich attenuator. In comparison, if region 3 forms a stem-loop with region 2, transcription will not be terminated because a 3–4 stem-loop cannot form.

Conditions that favor the formation of the 3–4 stem-loop ultimately rely on the translation of the *trpL* gene. As shown in **Figure 16.13**, three scenarios are possible. In Figure 16.13a, translation is not coupled with transcription. Region 1 rapidly hydrogen bonds to region 2, and region 3 is left to hydrogen bond to region 4. Therefore, the terminator stem-loop forms, and transcription is terminated just past the *trpL* gene at the U-rich attenuator.

In Figure 16.13b, coupled transcription and translation occur under conditions in which the tryptophan concentration is low. When tryptophan levels are low, the cell cannot make a sufficient amount of charged tRNA^{Trp}. As we see in Figure 16.13b, the ribosome pauses at the Trp codons in the *trpL* mRNA because it is waiting for charged tRNA^{Trp}. This pause occurs in such a way that the ribosome shields region 1 of the mRNA. This prevents region 1 from hydrogen bonding to region 2. As an alternative, region 2 hydrogen bonds to region 3. Therefore, because region 3 is already hydrogen bonded to region 2, the 3–4 stem-loop cannot form. Under these conditions, transcriptional termination does not occur, and RNA polymerase transcribes the rest of the operon. This ultimately enables the bacterium to make more tryptophan.

Finally, in Figure 16.13c, coupled transcription and translation occur under conditions in which a sufficient amount of tryptophan is present in the cell. In this case, translation of the *trpL* mRNA progresses to its stop codon, where the ribosome

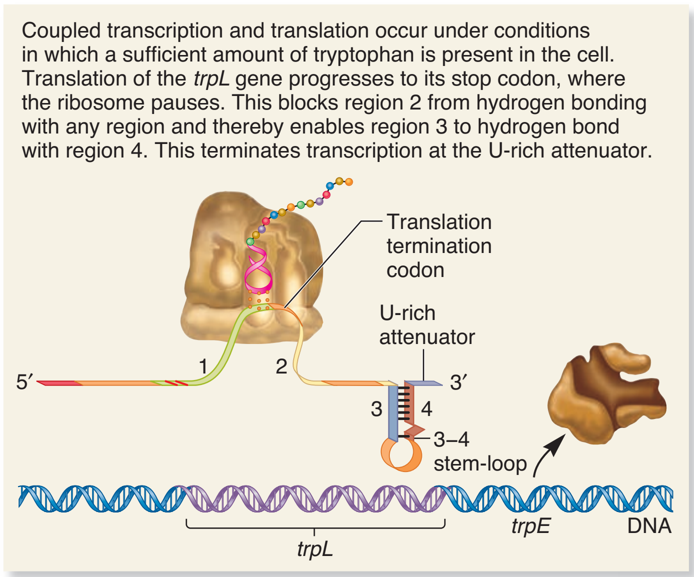


When translation is not coupled with transcription, the most stable form of the mRNA occurs when region 1 hydrogen bonds to region 2 and region 3 hydrogen bonds to region 4. A terminator stem-loop forms, and transcription will be terminated just past the *trpL* gene.

(a) No translation

Coupled transcription and translation occur under conditions in which the tryptophan concentration is very low. The ribosome pauses at the Trp codons in the *trpL* gene because insufficient amounts of charged tRNA^{Trp} are present. This pause blocks region 1 of the mRNA, so region 2 can hydrogen bond only with region 3. When this happens, the 3–4 stem-loop cannot form. Transcriptional termination does not occur, and RNA polymerase transcribes the rest of the operon.

(b) Low tryptophan levels, 2–3 stem-loop forms



Coupled transcription and translation occur under conditions in which a sufficient amount of tryptophan is present in the cell. Translation of the *trpL* gene progresses to its stop codon, where the ribosome pauses. This blocks region 2 from hydrogen bonding with any region and thereby enables region 3 to hydrogen bond with region 4. This terminates transcription at the U-rich attenuator.

(c) High tryptophan levels, 3–4 stem-loop forms

FIGURE 16.13 **Possible stem-loops formed from *trpL* mRNA under different conditions of translation.** Attenuation occurs in parts (a) and (c) due to the formation of a 3–4 stem-loop.

Concept check: Explain how the presence of tryptophan favors the formation of the 3–4 stem-loop.

pauses. The pausing at the stop codon prevents region 2 from hydrogen bonding with any region, thereby enabling region 3 to hydrogen bond with region 4. As in Figure 16.13a, this terminates transcription. Of course, keep in mind that the *trpL* mRNA contains two tryptophan codons. For the ribosome to smoothly progress to the *trpL* stop codon, enough charged tRNA^{Trp} must be available to translate this mRNA. It follows that the bacterium must have a sufficient amount of tryptophan. Under these conditions, the rest of the transcription of the operon is terminated.

Attenuation is a mechanism to regulate transcription that is found in several other operons involved with amino acid biosynthesis. In all cases, the mRNAs that encode the leader peptides are rich in codons for the particular amino acid that is synthesized by the enzymes encoded by the particular operon. For example, the mRNA that encodes the leader peptide of the histidine operon has seven histidine codons in its sequence, and the mRNA for the leader peptide of the leucine operon has four leucine codons. Like the *trp* operon, these other operons have alternative stem-loops, one of which is a transcriptional terminator.

Inducible Operons Encode Catabolic Enzymes, and Repressible Operons Usually Encode Anabolic Enzymes

Thus far, we have seen that bacterial genes can be transcriptionally regulated in a positive or negative way—and sometimes both. The *lac* operon is an inducible system regulated by sugar molecules that activate transcription. By comparison, the *trp* operon is a repressible operon regulated by tryptophan, a corepressor that binds to the repressor and turns the operon off. In addition, an abundance of charged tRNA^{Trp} in the cytoplasm can turn the *trp* operon off via attenuation.

By studying the genetic regulation of many operons, geneticists have discovered a general trend concerning inducible versus repressible regulation. When the genes in an operon encode proteins that function in the catabolism or breakdown of a substance, they are usually regulated in an inducible manner. The substance to be broken down or a related compound often acts as the inducer. For example, allolactose acts as an inducer of the *lac* operon. An inducible form of regulation allows the bacterium to express the appropriate genes only when they are needed to catabolize lactose.

In contrast, other enzymes are important for the anabolism or synthesis of small molecules. The genes that encode these anabolic enzymes tend to be regulated by a repressible mechanism. The corepressor or inhibitor is commonly the small molecule that is the product of the enzymes' biosynthetic activities. For example, tryptophan is produced by the sequential action of several enzymes that are encoded by the *trp* operon. Tryptophan itself acts as a corepressor that binds to the trp repressor protein when the intracellular levels of tryptophan become relatively high. This mechanism turns off the genes required for tryptophan biosynthesis when enough of this amino acid has been made. Therefore, genetic regulation via repression provides the bacterium with a way to prevent the overproduction of the product of a biosynthetic pathway.

16.3 REVIEWING THE KEY CONCEPTS

- The *trp* operon is regulated by the trp repressor that binds to the *trp* operator when tryptophan, a corepressor, is present (see Figure 16.11).
- A second way that the *trp* operon is regulated is via attenuation in which the formation of a terminator stem-loop causes early termination of transcription (see Figures 16.12, 16.13).
- Inducible operons typically encode catabolic enzymes, whereas repressible operons often encode anabolic enzymes.

16.3 COMPREHENSION QUESTIONS

1. When tryptophan binds to the trp repressor, this causes the trp repressor to __________ to the *trp* operator and ________ transcription.
   a. bind, inhibit
   b. not bind, inhibit
   c. bind, activate
   d. not bind, activate
2. During attenuation, when tryptophan levels are high, the ________ stem-loop forms and transcription _________ the *trpL* gene.
   a. 1–2, ends just past
   b. 3–4, ends just past
   c. 1–2, continues beyond
   d. 3–4, continues beyond
3. Operons involved with the biosynthesis of molecules, such as amino acids, are most likely to be regulated in which of the following manner?
   a. The product of the biosynthetic pathway represses transcription.
   b. The product of the biosynthetic pathway activates transcription.
   c. A precursor of the biosynthetic pathway represses transcription.
   d. A precursor of the biosynthetic pathway activates transcription.

16.4 TRANSLATIONAL AND POSTTRANSLATIONAL REGULATION

Though genetic regulation in bacteria occurs predominantly at transcription, many examples are known in which regulation is exerted at a later stage in gene expression. In some cases, specialized mechanisms have evolved to regulate the translation of certain mRNAs. Recall that the translation of mRNA occurs in three stages: initiation, elongation, and termination. Genetic regulation of translation is usually aimed at preventing the initiation step. In addition, as described later in Section 16.5, translation can be regulated by riboswitches.

The net result of translation is the synthesis of a protein. The activities of proteins within living cells and organisms ultimately determine an individual's traits. Therefore, to fully understand how proteins influence an organism's traits, researchers have investigated how the functions of proteins are regulated. The term **posttranslational** regulation refers to the functional control of proteins that are already present in the cell rather than regulation of transcription or translation. Posttranslational control can either activate or inhibit the function of a protein. Compared with transcriptional or translational regulation, posttranslational control can be relatively fast, occurring in a matter of seconds, which is an important advantage. In contrast, transcriptional and translational regulation typically require several minutes or even hours to take effect because these two mechanisms involve the synthesis and turnover of mRNA and polypeptides. In this section, we will examine some of the ways that bacteria can regulate the initiation of translation, as well as ways that protein function can be regulated posttranslationally.

Repressor Proteins and Antisense RNA Can Inhibit Translation

For some bacterial genes, the translation of mRNA is regulated by the binding of proteins or other RNA molecules that influence the ability of ribosomes to translate the mRNA into a polypeptide. A **translational regulatory protein** recognizes sequences within the mRNA, much as transcription factors recognize DNA sequences. In most cases, translational regulatory proteins act to inhibit translation. These are known as **translational repressors.** When a translational repressor protein binds to the mRNA, it can inhibit translational initiation in one of two ways. One possibility is that it can bind in the vicinity of the Shine-Dalgarno sequence and/or the start codon, thereby blocking the ribosome's ability to initiate translation in this region. Alternatively, the repressor protein may bind outside the Shine-Dalgarno/start codon region but stabilize an mRNA secondary structure that prevents initiation. Translational repression is also a form of genetic regulation found in eukaryotic species, and we will consider specific examples in Chapter 17.

A second way to regulate translation is via the synthesis of **antisense RNA,** an RNA strand that is complementary to a strand of mRNA. (The mRNA strand has the same sequence as the DNA sense strand.) To understand this form of genetic regulation, let's consider a trait known as osmoregulation, the ability to control the amount of water inside the cell. Osmoregulation is essential for the survival of most bacteria. Because the solute concentrations in the external environment may rapidly change between hypotonic and hypertonic conditions, bacteria must have an osmoregulation mechanism to maintain their internal cell volume. Otherwise, bacteria would be susceptible to the harmful effects of lysis or shrinking.

In *E. coli,* an outer membrane protein encoded by the *ompF* gene is important in osmoregulation. At low osmolarity, the ompF protein is preferentially produced, whereas at high osmolarity, its synthesis is decreased. The expression of another gene, known as *micF,* is responsible for inhibiting the expression of the *ompF* gene at high osmolarity. The name *mic* stands for mRNA-interfering complementary RNA. As shown in **Figure 16.14**, the *micF* RNA is complementary to the *ompF* mRNA; it is an antisense strand of RNA. When the *micF* gene is transcribed, its RNA product binds to the *ompF* mRNA via hydrogen bonding between their complementary regions. The binding of the *micF* RNA to the *ompF* mRNA prevents the *ompF* mRNA from being translated. The RNA transcribed from the *micF* gene is called antisense RNA because it is complementary to the *ompF* mRNA, which is a sense strand of mRNA that encodes a protein. The *micF* RNA does not encode a protein.

Posttranslational Regulation Can Occur Via Feedback Inhibition and Covalent Modifications

Let's now turn our attention to ways that protein function is regulated posttranslationally. A common mechanism to regulate the activity of metabolic enzymes is **feedback inhibition.** The synthesis of many cellular molecules such as amino acids, vitamins, and nucleotides occurs via the action of a series of enzymes that convert precursor molecules to a particular product. During feedback inhibition, the final product in a metabolic pathway inhibits an enzyme that acts early in the pathway.

Figure 16.15 depicts feedback inhibition in a metabolic pathway. Enzyme 1 is an example of an **allosteric enzyme,** an enzyme that contains two different binding sites. (The lac repressor is an allosteric protein, but not an enzyme.) The catalytic site is responsible for the binding of the substrate and its conversion to intermediate 1. The second site is a regulatory or allosteric site. This site binds the final product of the metabolic pathway. When bound to the regulatory site, the final product inhibits the catalytic ability of enzyme 1.

To appreciate feedback inhibition at the cellular level, we can consider the relationship between the product concentration and the regulatory site on enzyme 1. As the final product is made within the cell, its concentration gradually increases. Once the final product concentration has reached a level that is similar to its affinity for enzyme 1, the final product is likely to bind to the regulatory site on enzyme 1 and inhibit its function. In this way, the net result is that the final product of a metabolic pathway inhibits the further synthesis of more product. Under these

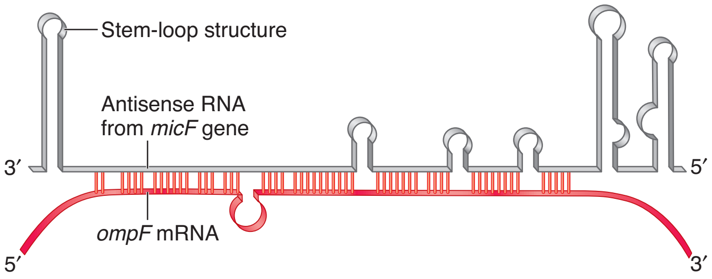
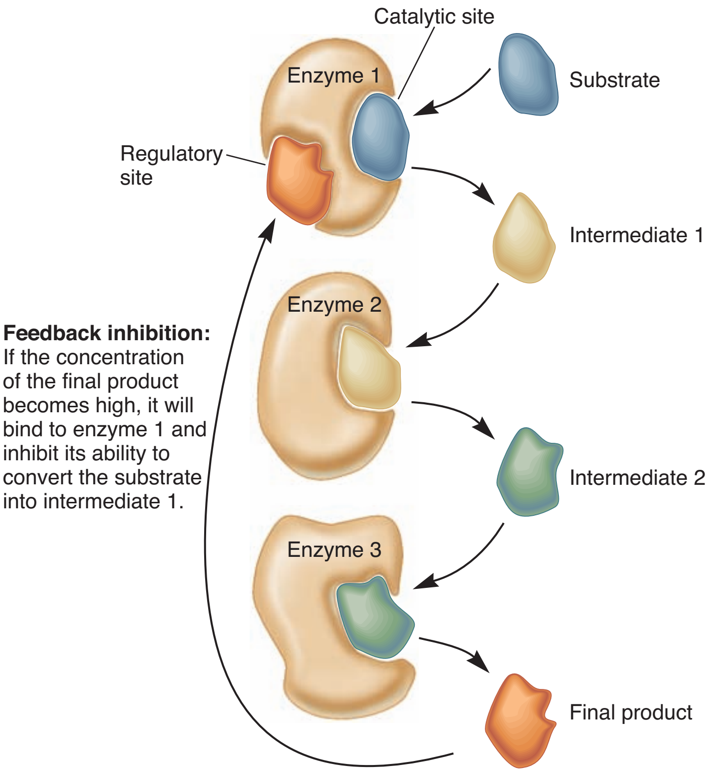


FIGURE 16.14 **The double-stranded RNA structure formed between the *micF* antisense RNA and the *ompF* mRNA.** Because they have regions that are complementary to each other, the *micF* antisense RNA hydrogen bonds to the *ompF* mRNA to form a double-stranded structure that prevents the *ompF* mRNA from being translated.

Concept check: How does micF antisense RNA affect the translation of ompF mRNA?



Feedback inhibition: If the concentration of the final product becomes high, it will bind to enzyme 1 and inhibit its ability to convert the substrate into intermediate 1.

FIGURE 16.15 **Feedback inhibition in a metabolic pathway.** The substrate is converted to a product by the sequential action of three different enzymes. Enzyme 1 has a catalytic site that recognizes the substrate; it also has a regulatory site that recognizes the final product. When the final product binds to the regulatory site, it inhibits the ability of enzyme 1 to convert the substrate into intermediate 1.

Concept check: Why is feedback inhibition an advantage to the bacterium?

conditions, the concentration of the final product has reached a level sufficient for the purpose of the bacterium.

A second strategy to control the function of proteins is by the covalent modification of their structure, a process called **posttranslational covalent modification.** Certain types of modifications are involved primarily in the assembly and construction of a functional protein (look ahead to Figure 22.3). These alterations include proteolytic processing; disulfide bond formation; and the attachment of prosthetic groups, sugars, or lipids. These are typically irreversible changes required to produce a functional protein. In contrast, other types of modifications, such as phosphorylation (—PO_4), acetylation (—$COCH_3$), and methylation (—CH_3), are often reversible modifications that transiently affect the function of a protein.

16.4 REVIEWING THE KEY CONCEPTS

- Translation can be regulated by translational regulatory proteins and by antisense RNA (see Figure 16.14).
- Posttranslational control of protein function may involve feedback inhibition and posttranslational covalent modifications (see Figure 16.15).

16.4 COMPREHENSION QUESTIONS

1. Translation can be regulated by
   a. translational repressors.
   b. antisense RNA.
   c. attenuation
   d. both a and b.
2. An example of a posttranslational covalent modification that may regulate protein function is
   a. phosphorylation.
   b. acetylation.
   c. methylation.
   d. all of the above.

16.5 RIBOSWITCHES

In 2001 and 2002, researchers in a few different laboratories discovered a mechanism of gene regulation called a **riboswitch.** In this form of regulation, an RNA molecule can exist in two different secondary conformations. One conformation is active, and the other inhibits gene expression. The conversion from one conformation to the other is due to the binding of a small molecule. As described in **Table 16.2**, a riboswitch can regulate transcription, translation, RNA stability, and splicing.

Riboswitches are widespread in bacteria. Researchers estimate that 3% to 5% of all bacterial genes may be regulated by riboswitches. In bacteria, they are involved in regulating genes associated with the biosynthesis of purines, amino acids, vitamins, and other essential molecules. Riboswitches are also found in archaea, algae, fungi, and plants. In this section, we will examine two examples of riboswitches in bacteria. The first example shows how riboswitches can regulate transcription and the second example involves translational regulation.

TABLE **16.2**

Types of Riboswitches

Type of Regulation	Description
Transcription	The 5′ region of an mRNA may exist in one conformation that forms a ρ-independent terminator, which causes attenuation of transcription. The other conformation does not form a terminator and is completely transcribed.
Translation	The 5′ region of an mRNA may exist in one conformation in which the Shine-Dalgarno sequence cannot be recognized by the ribosome, whereas the other conformation has an accessible Shine-Dalgarno sequence that allows the mRNA to be translated.
RNA stability	One mRNA conformation may be stable, whereas the other conformation acts as ribozyme that causes self-degradation.
Splicing	In eukaryotes, one pre-mRNA conformation may be spliced in one way, whereas another conformation is spliced in a different way.

A Riboswitch Can Regulate Transcription

Thiamin, also called vitamin B_1, is an important organic molecule for bacteria, archaea, and eukaryotes. The active form of this vitamin is thiamin pyrophosphate (TPP). TPP is an essential coenzyme for the functioning of a variety of enzymes such as certain enzymes in the citric acid cycle. In bacteria, TPP is made via biosynthetic enzymes that are encoded in the bacterial genome. For example, in *Bacillus subtilis*, the majority of genes involved in TPP synthesis are found within the *thi* operon that contains seven genes.

Certain genes that encode TPP biosynthetic enzymes, such as those of the *thi* operon in *B. subtilis*, are regulated by a riboswitch that controls transcription (**Figure 16.16**). As the polycistronic mRNA for the *thi* operon is being made, the 5′ end quickly folds into a secondary structure. When TPP levels are low, the secondary structure has a stem-loop called an antiterminator, which prevents the formation of the terminator stem-loop. Under these conditions, transcription of the entire *thi* operon occurs. In this way, the bacterium is able to make more TPP, which is in short supply. By comparison, when TPP levels are high, TPP binds to the RNA and causes a change in its secondary structure. As shown on the right side of Figure 16.16, a ρ-independent terminator forms instead of the antiterminator stem-loop. The ρ-independent terminator abruptly stops transcription, thereby inhibiting the production of the enzymes that are needed to make more TPP. Similar to the *trp* operon discussed earlier in this chapter, this is an example of attenuation.

A Riboswitch Can Regulate Translation

In gram-positive bacteria, such as *B. subtilis*, the regulation of TPP biosynthetic enzymes occurs via a riboswitch that controls transcription. However, in gram-negative bacteria, such as *E. coli*,

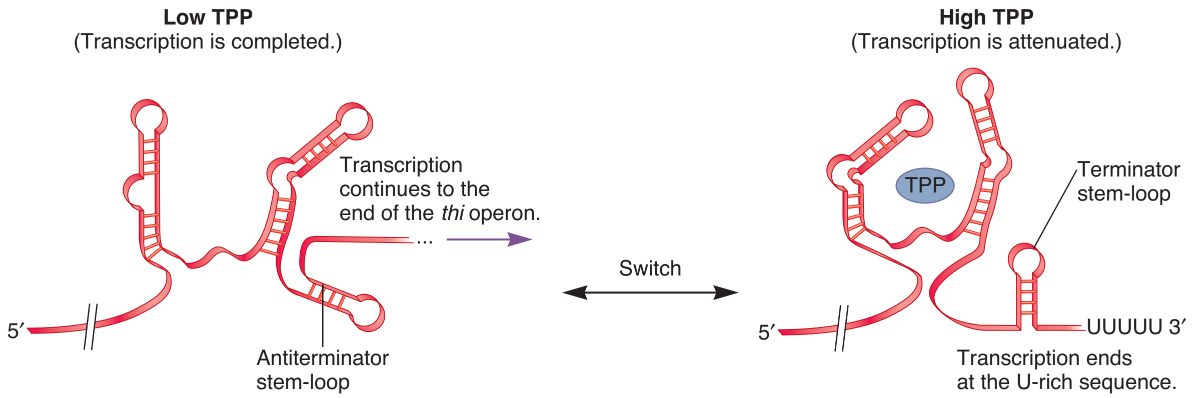

Low TPP
(Transcription is completed.)
Transcription continues to the end of the *thi* operon.
Antiterminator stem-loop
5′
Switch
High TPP
(Transcription is attenuated.)
TPP
Terminator stem-loop
5′
UUUUU 3′
Transcription ends at the U-rich sequence.

FIGURE 16.16 **A TPP riboswitch in *Bacillus subtilis* that regulates transcription.**

Concept check: Which RNA conformation favors transcription—the form with the antiterminator stem-loop or the form with the terminator stem-loop?

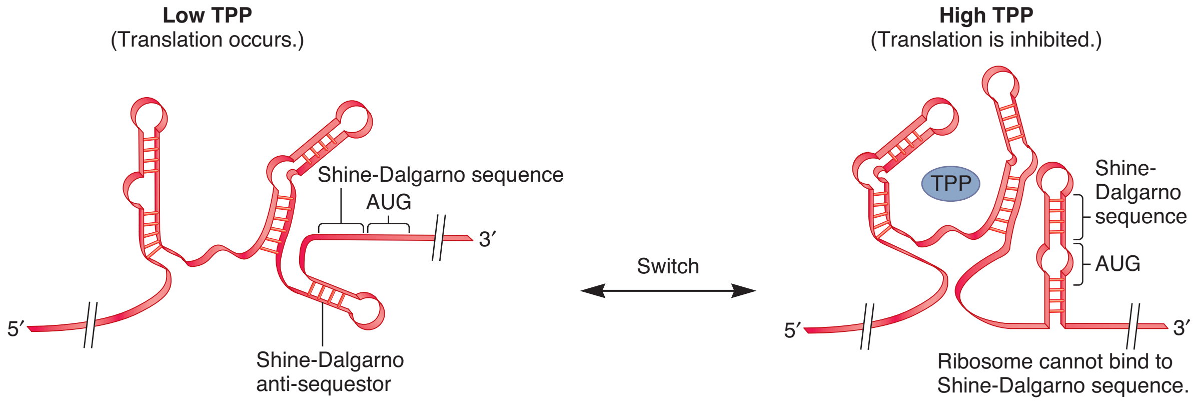

Low TPP
(Translation occurs.)
Shine-Dalgarno sequence
AUG
3′
5′
Shine-Dalgarno anti-sequestor
Switch
High TPP
(Translation is inhibited.)
TPP
Shine-Dalgarno sequence
AUG
5′
3′
Ribosome cannot bind to Shine-Dalgarno sequence.

FIGURE 16.17 **A TPP riboswitch in *E. coli* that regulates translation.** Note: In both cases, the *thiMD* operon is completely transcribed.

Concept check: Which RNA conformation favors translation—the form with the Shine-Dalgarno antisequestor or the form in which the Shine-Dalgarno sequence is within a stem-loop?

a structurally similar riboswitch regulates translation. **Figure 16.17** shows how a riboswitch can regulate translation. In *E. coli*, the *thiMD* operon encodes two enzymes involved with TPP biosynthesis. Recall from Chapter 15 that the Shine-Dalgarno sequence in an mRNA causes the mRNA to bind to the ribosome (refer back to Figure 15.17). When TPP levels are low, the 5′ end of the mRNA folds into a structure that contains a stem-loop called the Shine-Dalgarno anti-sequestor. When this stem-loop forms, the Shine-Dalgarno sequence is accessible, which allows the mRNA to bind to the ribosome. Therefore, the mRNA is translated when TPP is in short supply. By comparison, when TPP levels are high, TPP binds to the RNA and causes a change in its secondary structure. As shown on the right side of Figure 16.17, a stem-loop forms that contains the Shine-Dalgarno sequence and the start codon. The formation of this stem-loop sequesters the Shine-Dalgarno sequence, thereby preventing ribosomal binding. This blocks the translation of the enzymes that are needed to make more TPP.

When comparing Figures 16.16 and 16.17, it is worth noting that the 5′ end of the *thiMD* mRNA of *E. coli* can exist in two different secondary structures that greatly resemble those that occur at the 5′ end of the *thi* operon of *B. subtilis*. However, the effects on regulation are quite different. The TPP riboswitch in *E. coli* controls translation, whereas the TPP riboswitch in *B. subtilis* controls transcription.

16.5 REVIEWING THE KEY CONCEPTS

- A riboswitch is a form of regulation in which an RNA exists in two different secondary structures. The conversion from one secondary structure to the other is due to the binding of a small molecule.
- A riboswitch can regulate transcription, translation, RNA stability, and splicing (see Table 16.2, Figures 16.16, 16.17).

16.5 COMPREHENSION QUESTIONS

1. For a riboswitch that controls transcription, the binding of a small molecule such as TPP controls whether the RNA
   a. has an antiterminator or terminator stem-loop.
   b. has a Shine-Dalgarno anti-sequestor or the Shine-Dalgarno sequence in a stem-loop.
   c. is degraded from its 5′ end.
   d. both a and b.

KEY TERMS

Page 355. gene regulation, constitutive genes

Page 356. repressor, activator, negative control, positive control, inducer, inducible genes, corepressor, inhibitor, repressible genes

Page 358. enzyme adaptation, operon, polycistronic mRNA, promoter, terminator, CAP site, catabolite activator protein (CAP), operator site, operator, lac repressor

Page 359. induced, allosteric regulation, allosteric site

Page 362. merozygote

Page 364. *trans*-effect, *trans*-acting factor, *cis*-acting element, *cis*-effect, catabolite repression, diauxic growth, cyclic-AMP (cAMP)

Page 367. attenuation

Page 368. trp repressor, attenuator sequence

Page 371. posttranslational, translational regulatory protein, translational repressors

Page 372. antisense RNA, feedback inhibition, allosteric enzyme

Page 373. posttranslational covalent modification, riboswitch

CHAPTER SUMMARY

- Gene regulation is the phenomenon in which the level of gene expression can vary under different conditions. By comparison, constitutive genes are expressed at relatively constant levels. Gene regulation can occur at transcription, translation, or posttranslationally (see Figure 16.1).

16.1 Overview of Transcriptional Regulation

- Repressors and activators are proteins that exert negative control or positive control, respectively. Small effector molecules control the function of repressors and activators. These include inducers, inhibitors, and corepressors (see Figure 16.2).

16.2 Regulation of the *lac* Operon

- Enzyme adaptation is the phenomenon in which the enzymes that metabolize a molecule are made only when the molecule is present in the environment.
- The *lac* operon encodes a polycistronic mRNA for proteins that are involved with the uptake and metabolism of lactose (see Figure 16.3).
- The lac repressor protein binds to the *lac* operator and inhibits transcription. Allolactose binds to the repressor and causes a conformational change that prevents the repressor from binding to the operator. This event induces transcription (see Figure 16.4).
- The regulation of the *lac* operon enables the bacterium *E. coli* to respond to changes in the level of lactose in its environment (see Figure 16.5).
- Jacob and Monod constructed merozygotes to show that the *lacI* gene encodes a diffusible repressor protein (see Figures 16.6, 16.7).
- Mutations in the *lacI* gene and the *lac* operator site may have different effects in a merozygote (see Table 16.1).
- CAP is an activator protein of the *lac* operon. It binds to the CAP site when cAMP levels are high. Glucose inhibits cAMP levels, thereby inhibiting the *lac* operon (see Figure 16.8).
- The lac repressor, which is a tetramer, binds to two operator sites, either to O_1 and O_2, or to O_1 and O_3, causing a loop to form in the DNA (see Figures 16.9, 16.10).

16.3 Regulation of the *trp* Operon

- The *trp* operon is regulated by the trp repressor that binds to the *trp* operator when tryptophan, a corepressor, is present (see Figure 16.11).
- A second way that the *trp* operon is regulated is via attenuation in which the formation of a terminator stem-loop causes early termination of transcription (see Figures 16.12, 16.13).
- Inducible operons typically encode catabolic enzymes, whereas repressible operons often encode anabolic enzymes.

16.4 Translational and Posttranslational Regulation

- Translation can be regulated by translational regulatory proteins and by antisense RNA (see Figure 16.14).
- Posttranslational control of protein function may involve feedback inhibition and posttranslational covalent modifications (see Figure 16.15).

16.5 Riboswitches

- A riboswitch is a form of regulation in which an RNA exists in two different secondary structures. The conversion from one secondary structure to the other is due to the binding of a small molecule.
- A riboswitch can regulate transcription, translation, RNA stability, and splicing (see Table 16.2, Figures 16.16, 16.17).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Researchers have identified mutations in the promoter region of the *lacI* gene that make it more difficult for the *lac* operon to be induced. These are called *lacI*^{Q} mutants, because a greater Quantity of lac repressor protein is made. Explain why an increased transcription of the *lacI* gene makes it more difficult to induce the *lac* operon.

Answer: An increase in the amount of lac repressor protein makes it easier to repress the *lac* operon. When the cells become exposed to lactose, allolactose levels slowly rise. Some of the allolactose binds to the lac repressor protein and causes it to be released from the operator site. If many more lac repressor proteins are found within the cell, more allolactose is needed to ensure that no unoccupied repressor proteins can repress the operon.

S2. Explain how the pausing of the ribosome in the presence or absence of tryptophan affects the formation of a terminator stem-loop.

Answer: The key issue is the location where the ribosome stalls. In the absence of tryptophan, it stalls over the Trp codons in the *trpL* mRNA. Stalling at this site shields region 1 in the attenuator region. Because region 1 is unavailable to hydrogen bond with region 2, region 2 hydrogen bonds with region 3. Therefore, region 3 cannot form a terminator stem-loop with region 4. Alternatively, if tryptophan levels in the cell are sufficient, the ribosome pauses over the stop codon in the *trpL* RNA. In this case, the ribosome shields region 2. Therefore, regions 3 and 4 hydrogen bond with each other to form a terminator stem-loop, which abruptly halts the continued transcription of the *trp* operon.

S3. In bacteria, it is common for two or more genes to be arranged together in an operon. Discuss the arrangement of genetic sequences within an operon. What is the biological advantage of an operon organization?

Answer: An operon contains several different DNA sequences that play specific roles. For transcription to take place, an operon is flanked by a promoter to signal the beginning of transcription and a terminator to signal the end of transcription. Two or more genes are found between these two sequences. A key feature of an operon is that the expression of the genes occurs as a unit. When transcription takes place, a polycistronic RNA is made that encodes all of the genes.

In order to control the ability of RNA polymerase to transcribe an operon, an additional DNA sequence, known as the operator site, is usually present. The base sequence within the operator site can serve as a binding site for genetic regulatory proteins called activators or repressors. The advantage of an operon organization is that it allows a bacterium to coordinately regulate a group of genes with a common function. For example, an operon may contain a group of genes involved in lactose breakdown or a group of genes involved in tryptophan synthesis. The genes within an operon usually encode proteins within a common metabolic pathway or cellular function.

S4. The sequential use of two sugars by a bacterium is known as diauxic growth. It is a common phenomenon among many bacterial species. When glucose is one of the two sugars available, it is typical that the bacterium metabolizes glucose first, and then a second sugar after the glucose has been used up. Among *E. coli* and related species, diauxic growth is regulated by intracellular cAMP levels and the catabolite activator protein (CAP). Summarize the effects of glucose and lactose on the ability of the lac repressor and the cAMP-CAP complex to regulate the *lac* operon.

Answer: In the absence of lactose, the lac repressor has the dominating effect of shutting off the *lac* operon. Even when glucose is also absent and the cAMP-CAP complex is formed, the presence of the bound lac repressor prevents the expression of the *lac* operon. The effects of the cAMP-CAP complex are exerted only in the presence of lactose. When lactose is present and glucose is absent, the cAMP-CAP complex increases the rate of transcription. However, when both lactose and glucose are present, the inability of CAP to bind to the *lac* operon decreases the rate of transcription. The table shown here summarizes these effects.

***lac* Operon Regulation**

Sugar Present	Transcription of the *lac* Operon
None	The operon is turned off due to the dominating effect of the lac repressor protein.
Lactose	The operon is maximally turned on. The repressor protein is removed from the operator site, and the cAMP-CAP complex is bound to the CAP site.
Glucose	The operon is turned off due to the dominating effect of the lac repressor protein.
Lactose and glucose	The expression of the *lac* operon is greatly decreased. The lac repressor is removed from the operator site, and the catabolite activator protein is not bound to the CAP site. The absence of CAP at the CAP site makes it difficult for RNA polymerase to begin transcription. However, a little more transcription occurs under these conditions than in the absence of lactose, when the repressor is bound.

Conceptual Questions

C1. What is the difference between a constitutive gene and a regulated gene?

C2. In general, why is it important to regulate genes? Discuss examples of situations in which it would be advantageous for a bacterial cell to regulate genes.

C3. If a gene is repressible and under positive control, describe what kind of effector molecule and regulatory protein are involved. Explain how the binding of the effector molecule affects the regulatory protein.

C4. Transcriptional regulation often involves a regulatory protein that binds to a segment of DNA and a small effector molecule that binds to the regulatory protein. Do the following terms apply to a regulatory protein, a segment of DNA, or a small effector molecule?

A. Repressor
B. Inducer
C. Operator site
D. Corepressor
E. Activator
F. Attenuator
G. Inhibitor

C5. An operon is repressible—a small effector molecule turns off transcription. Which combinations of small effector molecules and regulatory proteins could be involved?

A. An inducer plus a repressor
B. A corepressor plus a repressor
C. An inhibitor plus an activator
D. An inducer plus an activator

C6. Some mutations have a *cis*-effect on gene expression, whereas others have a *trans*-effect. Explain the molecular differences between *cis*- and *trans*-mutations. Which type of mutation (*cis* or *trans*) can be complemented in a merozygote experiment?

C7. What is enzyme adaptation? From a genetic point of view, how does it occur?

C8. In the *lac* operon, how would gene expression be affected if one of the following segments was missing?

A. *lac* operon promoter

B. Operator site

C. *lacA* gene

C9. If an abnormal repressor protein could still bind allolactose, but the binding of allolactose did not alter the conformation of the repressor protein, how would this affect the expression of the *lac* operon?

C10. What is diauxic growth? Explain the roles of cAMP and the catabolite activator protein (CAP) in this process.

C11. Mutations may have an effect on the expression of the *lac* operon and the *trp* operon. Would the following mutations have a *cis*- or *trans*-effect on the expression of the structural genes in the operon?

A. A mutation in the operator site that prevents the lac repressor from binding to it

B. A mutation in the *lacI* gene that prevents the lac repressor from binding to DNA

C. A mutation in *trpL* that prevents attenuation

C12. Would a mutation that inactivated the lac repressor and prevented it from binding to the *lac* operator site result in the constitutive expression of the *lac* operon under all conditions? Explain. What is the disadvantage to the bacterium of having a constitutive *lac* operon?

C13. What is meant by the term *attenuation*? Is it an example of gene regulation at the level of transcription or translation? Explain your answer.

C14. As described in Figure 16.12, four regions within the *trpL* mRNA can form stem-loops. Let's suppose that mutations have been previously identified that prevent the ability of a particular region to form a stem-loop with a complementary region. For example, a region 1 mutant cannot form a 1–2 stem-loop, but it can still form a 2–3 or 3–4 stem-loop. Likewise, a region 4 mutant can form a 1–2 or 2–3 stem-loop but not a 3–4 stem-loop. Under the following conditions, would attenuation occur?

A. Region 1 is mutant, tryptophan is high, and translation is not occurring.

B. Region 2 is mutant, tryptophan is low, and translation is occurring.

C. Region 3 is mutant, tryptophan is high, and translation is not occurring.

D. Region 4 is mutant, tryptophan is low, and translation is not occurring.

C15. As described in Chapter 15, enzymes known as aminoacyl-tRNA synthetases are responsible for attaching amino acids to tRNAs. Let's suppose that tryptophanyl-tRNA synthetase was partially defective at attaching tryptophan to tRNA; its activity was only 10% of that found in a normal bacterium. How would that affect attenuation of the *trp* operon? Would it be more or less likely to be attenuated? Explain your answer.

C16. The 3–4 stem-loop and U-rich attenuator found in the *trp* operon (see Figure 16.12) is an example of ρ-independent termination. The mechanism of ρ-independent termination is described in Chapter 14. Would you expect attenuation to occur if the tryptophan levels were high and mutations changed the attenuator sequence from UUUUUUUU to UGGUUGUC? Explain why or why not.

C17. Mutations in tRNA genes can create tRNAs that recognize stop codons. Because stop codons are sometimes called nonsense codons, these types of mutations that affect tRNAs are called nonsense suppressors. For example, a normal tRNA^{Gly} has an anticodon sequence CCU that recognizes a glycine codon in mRNA (GGA) and puts in a glycine during translation. However, a mutation in the gene that encodes tRNA^{Gly} could change the anticodon to ACU. This mutant tRNA^{Gly} would still carry glycine, but it would recognize the stop codon UGA. Would this mutation affect attenuation of the *trp* operon? Explain why or why not. Note: To answer this question, you need to look carefully at Figure 16.12 and see if you can identify any stop codons that may exist beyond the UGA stop codon that is found after region 1.

C18. Translational control is usually aimed at preventing the initiation of translation. With regard to cellular efficiency, why do you think this is the case?

C19. What is antisense RNA? How does it affect the translation of a complementary mRNA?

C20. A species of bacteria can synthesize the amino acid histidine so it does not require histidine in its growth medium. A key enzyme, which we will call histidine synthetase, is necessary for histidine biosynthesis. When these bacteria are given histidine in their growth medium, they stop synthesizing histidine intracellularly. Based on this observation alone, propose three different regulatory mechanisms to explain why histidine biosynthesis ceases when histidine is in the growth medium. To explore this phenomenon further, you measure the amount of intracellular histidine synthetase protein when cells are grown in the presence and absence of histidine. In both conditions, the amount of this protein is identical. Which mechanism of regulation would be consistent with this observation?

C21. Using three examples, describe how allosteric sites are important in the function of genetic regulatory proteins.

C22. In what ways are the actions of the lac repressor and trp repressor similar and how are they different with regard to their binding to operator sites, their effects on transcription, and the influences of small effector molecules?

C23. Transcriptional repressor proteins (e.g., lac repressor), antisense RNA, and feedback inhibition are three different mechanisms that turn off the expression of genes and gene products. Which of these three mechanisms would be most effective in each of the following situations?

A. Shutting down the synthesis of a polypeptide

B. Shutting down the synthesis of mRNA

C. Shutting off the function of a protein

For your answers in parts A–C that have more than one mechanism, which mechanism is the fastest or the most efficient?

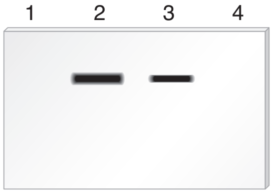
Application and Experimental Questions

E1. Answer the following questions that pertain to the experiment of Figure 16.7.

A. Why was β-ONPG used? Why was no yellow color observed in one of the four tubes? Can you propose alternative methods for measuring the level of expression of the *lac* operon?

B. The optical density values were twice as high for the mated strain as for the parent strain. Why was this result obtained?

E2. Chapter 19 describes a blotting method known as Northern blotting, which can be used to detect RNA transcribed from a particular gene or a particular operon. In this method, a specific RNA is detected by using a short segment of cloned DNA as a probe. The DNA probe, which is radioactive, is complementary to the RNA that the researcher wishes to detect. After the radioactive probe DNA binds to the RNA within a blot of a gel, the RNA is visualized as a dark band on an X-ray film. For example, a DNA probe complementary to the mRNA of the *lac* operon could be used to specifically detect the *lac* operon mRNA on a gel blot. As shown here, the method of Northern blotting can be used to determine the amount of a particular RNA transcribed under different types of growth conditions. In this Northern blot, bacteria containing a normal *lac* operon were grown under different types of conditions, and then the mRNA was isolated from the cells and subjected to Northern blotting, using a probe that is complementary to the mRNA of the *lac* operon.



Lane 1. Growth in medium containing glucose
Lane 2. Growth in medium containing lactose
Lane 3. Growth in medium containing glucose and lactose
Lane 4. Growth in medium that doesn’t contain glucose or lactose

Based on your understanding of the regulation of the *lac* operon, explain these results. Which is more effective at shutting down the *lac* operon, the binding of the *lac* repressor or the removal of CAP? Explain your answer based on the results shown in the Northern blot.

E3. As described in experimental question E2 and also in Chapter 19, the technique of Northern blotting can be used to detect the transcription of RNA. Draw the results you would expect from a Northern blot if bacteria were grown in a medium containing lactose (and no glucose) but had the following mutations:

Lane 1. Normal strain

Lane 2. Strain with a mutation that inactivates the lac repressor

Lane 3. Strain with a mutation that prevents allolactose from binding to the lac repressor

Lane 4. Strain with a mutation that inactivates CAP

How would your results differ if these bacterial strains were grown in media that did not contain lactose or glucose?

E4. An absentminded researcher follows the protocol described in Figure 16.7 and (at the end of the experiment) does not observe any yellow color in any of the tubes. Yikes! Which of the following mistakes could account for this observation?

A. Forgot to sonicate the cells

B. Forgot to add lactose to two of the tubes

C. Forgot to add β-ONPG to the four tubes

E5. Explain how the data shown in Figure 16.9 indicate that two operator sites are necessary for repression of the *lac* operon. What would the results have been if all three operator sites were required for the binding of the lac repressor?

E6. A mutant strain has a defective *lac* operator site that results in the constitutive expression of the *lac* operon. Outline an experiment you would carry out to demonstrate that the operator site must be physically adjacent to the genes that it influences. Based on your knowledge of the *lac* operon, describe the results you would expect.

E7. Let’s suppose you have isolated a mutant strain of *E. coli* in which the *lac* operon is constitutively expressed. To understand the nature of this defect, you create a merozygote in which the mutant strain contains an F′ factor with a normal *lac* operon and a normal *lacI* gene. You then compare the mutant strain and the merozygote with regard to their β-galactosidase activities in the presence and absence of lactose. You obtain the following results:

	Addition of Lactose	**Amount of β-Galactosidase (percentage of mutant strain in the presence of lactose)**
Mutant	No	100
Mutant	Yes	100
Merozygote	No	100
Merozygote	Yes	200

Explain the nature of the defect in the mutant strain.

Questions for Student Discussion/Collaboration

1. Discuss the advantages and disadvantages of genetic regulation at the different levels described in Figure 16.1.
2. As you look at Figure 16.10, discuss possible “molecular ways” that the cAMP-CAP complex and lac repressor may influence RNA polymerase function. In other words, try to explain how the bending and looping in DNA may affect the ability of RNA polymerase to initiate transcription.

Answers to Comprehension Questions

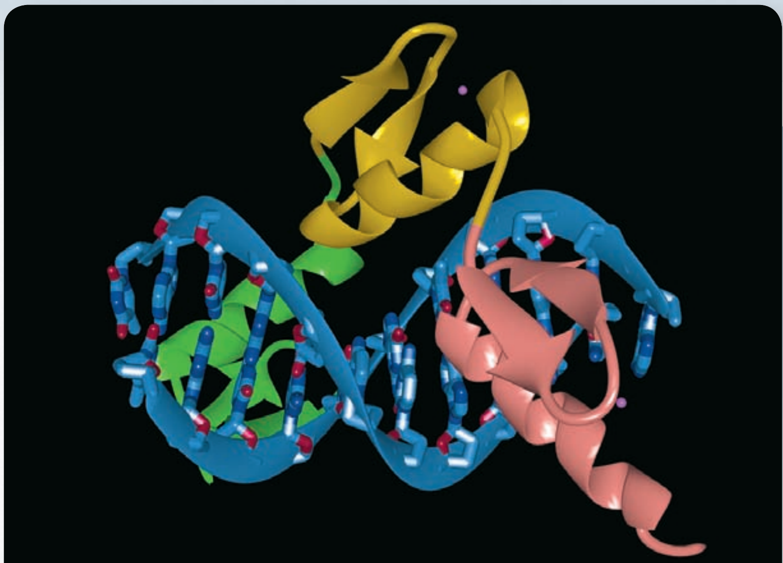
16.1: c, a
16.2: b, d, b, d
16.3: a, b, a
16.4: d, d
16.5: a

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

17



Certain proteins, known as regulatory transcription factors, have the ability to bind into the major groove of DNA and regulate gene transcription.

CHAPTER OUTLINE

17.1 Regulatory Transcription Factors

17.2 Chromatin Remodeling, Histone Variation, and Histone Modification

17.3 DNA Methylation

17.4 Insulators

17.5 Regulation of RNA Processing, RNA Stability, and Translation

GENE REGULATION IN EUKARYOTES

As we learned in Chapter 16, **gene regulation** enables bacterial cells to control the level of gene expression so that genes are expressed at high or low levels. Likewise, the ability to regulate genes provides many benefits to eukaryotic organisms, a category that includes protists, fungi, plants, and animals. Like their prokaryotic counterparts, eukaryotic cells need to adapt to changes in their environment. For example, eukaryotic cells respond to changes in nutrient availability by enzyme adaptation, much as prokaryotic cells do. Eukaryotic cells also respond to environmental stresses such as ultraviolet (UV) radiation by inducing genes that provide protection against harmful environmental agents. An example is the ability of humans to develop a tan. The tanning response helps to protect a person's skin cells against the damaging effects of UV rays.

Among plants and animals, multicellularity and a more complex cell structure also demand a much greater level of gene regulation. The life cycle of complex eukaryotic organisms involves the progression through developmental stages that lead to a mature organism. Some genes are expressed only during early stages of development, such as the embryonic stage, whereas others are expressed in the adult. In addition, complex eukaryotic species are composed of many different tissues that contain a variety of cell types. Gene regulation is necessary to ensure the differences in structure and function among distinct cell types. It is amazing that the various cells within a multicellular organism usually contain the same genetic material, yet phenotypically may look quite different. For example, the appearance of a human nerve cell seems about as similar to a muscle cell as an amoeba is to a paramecium. In spite of these phenotypic differences, a human nerve cell and muscle cell actually contain the same complement of human chromosomes. Nerve and muscle cells look strikingly different because of gene regulation rather than differences in DNA content. Many genes are expressed in the nerve cell and not the muscle cell, and vice versa.

The molecular mechanisms that underlie gene regulation in eukaryotes bear many similarities to the ways that bacteria regulate their genes. As in bacteria, regulation in eukaryotes can occur at any step in the pathway of gene expression (**Figure 17.1**). We will begin this chapter with an exploration of gene regulation at the level of transcription, an important form of control. In addition, research in the past few decades has revealed that eukaryotic organisms frequently regulate gene expression at points other than transcription. We will discuss some well-studied examples in which genes are regulated at these other control points.

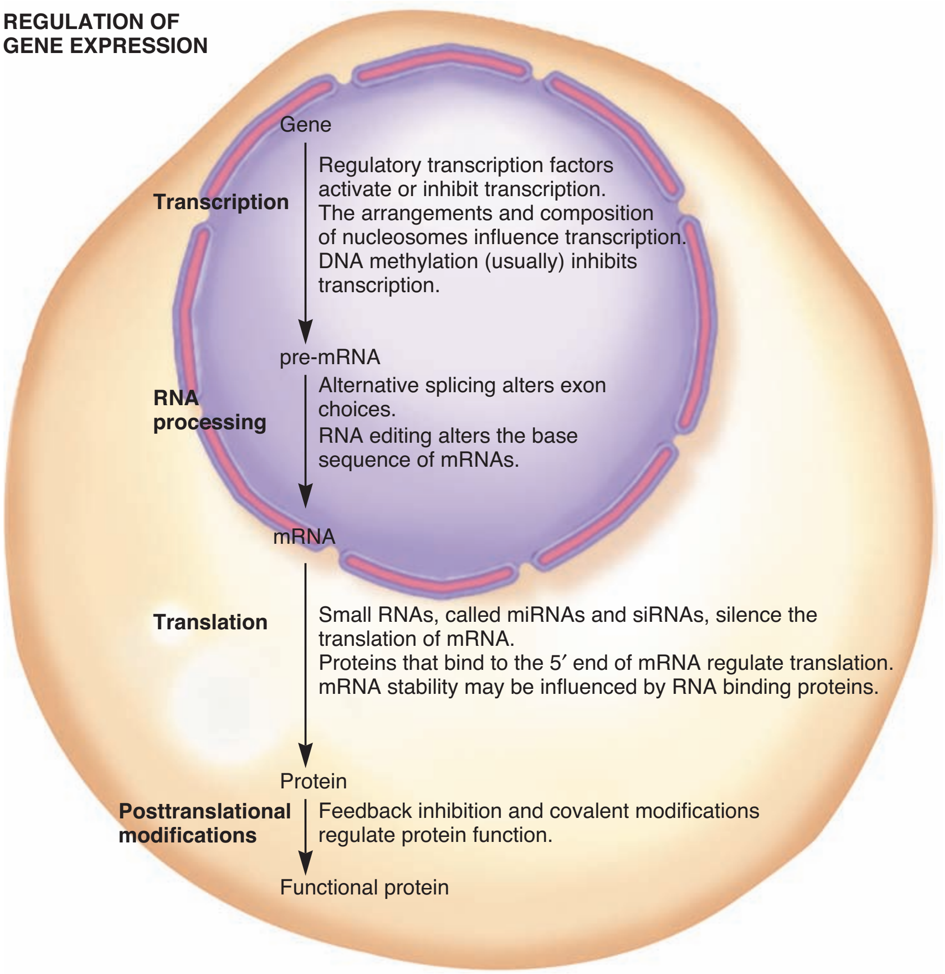

REGULATION OF GENE EXPRESSION
Gene
Transcription
Regulatory transcription factors activate or inhibit transcription.
The arrangements and composition of nucleosomes influence transcription.
DNA methylation (usually) inhibits transcription.
pre-mRNA
RNA processing
Alternative splicing alters exon choices.
RNA editing alters the base sequence of mRNAs.
mRNA
Translation
Small RNAs, called miRNAs and siRNAs, silence the translation of mRNA.
Proteins that bind to the 5′ end of mRNA regulate translation.
mRNA stability may be influenced by RNA binding proteins.
Protein
Posttranslational modifications
Feedback inhibition and covalent modifications regulate protein function.
Functional protein

FIGURE 17.1 **Levels of gene expression commonly subject to regulation.**

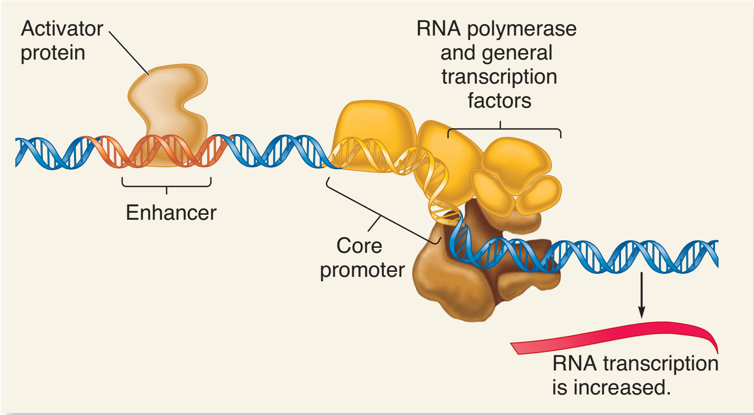
Concept check: Which of these levels is the most energy-efficient way to regulate gene expression?

17.1 REGULATORY TRANSCRIPTION FACTORS

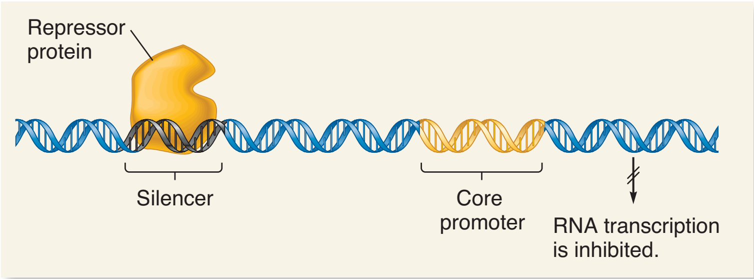
The term **transcription factor** is broadly used to describe proteins that influence the ability of RNA polymerase to transcribe a given gene. We will focus our attention on transcription factors that affect the ability of RNA polymerase to begin the transcription process. Such transcription factors can regulate the binding of the transcriptional apparatus to the core promoter and/or control the switch from the initiation to the elongation stage of transcription. Two categories of transcription factors play a key role in these processes. In Chapter 14, we considered **general transcription factors (GTFs),** which are required for the binding of RNA polymerase to the core promoter and its progression to the elongation stage. General transcription factors are necessary for a basal level of transcription. In addition, eukaryotic cells possess a diverse array of **regulatory transcription factors** that serve to regulate the rate of transcription of target genes. The importance of transcription factors is underscored by the large number of genes that encode this category of proteins. A sizable portion of the genomes of complex organisms such as plants and animals is devoted to the process of gene regulation. For example, in *Arabidopsis thaliana*, a plant that is used as a model organism by plant geneticists, approximately 5% of its genome encodes transcription factors. This species has more than 1500 different genes that encode proteins that regulate the transcription of other genes. Similarly, 2% to 3% of the genes in the human genome encode transcription factors.

As discussed in this section, regulatory transcription factors exert their effects by influencing the ability of RNA polymerase to begin transcription of a particular gene. They recognize *cis*-acting elements that are often located in the vicinity of the core promoter but may be relatively far away. These DNA sequences are analogous to the operator sites that control bacterial promoters. In eukaryotes, these DNA sequences are generally known as **regulatory elements,** or **control elements.** When a regulatory transcription factor binds to a regulatory element, it affects the transcription of an associated gene. For example, the binding of regulatory transcription factors may enhance the rate of transcription (**Figure 17.2a**). Such a transcription factor is termed an **activator,** and the sequence it binds to is called an **enhancer.** An enhancer may be recognized by more than one type of activator. Alternatively, regulatory transcription factors may act as **repressors** by binding to elements called **silencers** and preventing transcription from occurring (**Figure 17.2b**).

By studying transcriptional regulation, researchers have discovered that most eukaryotic genes, particularly those found in multicellular species, are regulated by many factors. This phenomenon is called **combinatorial control** because the combination of many factors determines the expression of any given


Activator protein
RNA polymerase and general transcription factors
Enhancer
Core promoter
RNA transcription is increased.

(a) Gene activation



(b) Gene repression

FIGURE 17.2 **Overview of transcriptional regulation by regulatory transcription factors.** These proteins can act as either **(a)** an activator to increase the rate of transcription or **(b)** a repressor to decrease the rate of transcription.

gene. At the level of transcription, the following are common factors that contribute to combinatorial control:

1. One or more activator proteins may stimulate the ability of RNA polymerase to initiate transcription.
2. One or more repressor proteins may inhibit the ability of RNA polymerase to initiate transcription.
3. The function of activators and repressors may be modulated in a variety of ways, including the binding of small effector molecules, protein-protein interactions, and covalent modifications.
4. Regulatory proteins may alter the composition or arrangements of nucleosomes in the vicinity of a promoter, thereby affecting transcription.
5. DNA methylation may inhibit transcription, either by preventing the binding of an activator protein or by recruiting proteins that cause the chromatin to become more compact.

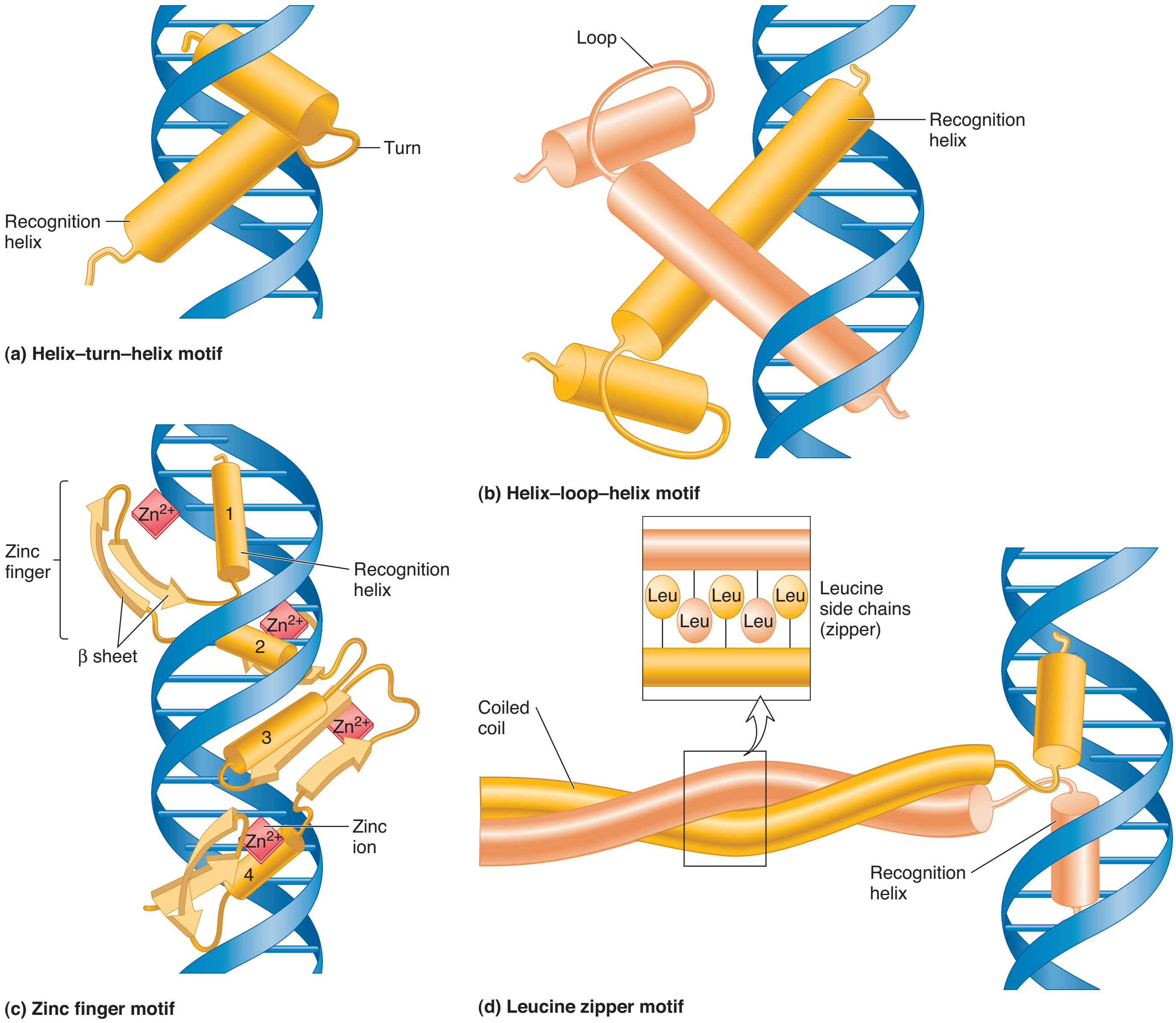
All five of these factors can contribute to the regulation of a single gene, or possibly only three or four will play a role. In most cases, transcriptional regulation is aimed at controlling the initiation of transcription at the promoter. In this section, we will survey the first three factors that affect transcription. In the sections to follow, we will consider the fourth and fifth factors.

Structural Features of Regulatory Transcription Factors Allow Them to Bind to DNA

Genes that encode general and regulatory transcription factor proteins have been identified and sequenced from a wide variety of eukaryotic species, including yeast, plants, and animals. Several different families of evolutionarily related transcription factors have been discovered. In recent years, the molecular structures of transcription factor proteins have become an area of intense research. Transcription factor proteins contain regions, called **domains,** that have specific functions. For example, one domain of a transcription factor may have a DNA-binding function, and another may provide a binding site for a small effector molecule. When a domain or portion of a domain has a very similar structure in many different proteins, such a structure is called a **motif.**

Figure 17.3 depicts several different domain structures found in transcription factor proteins. The protein secondary structure known as an α helix is frequently found in transcription factors. Why is the α helix common in such proteins? The explanation is that the α helix is the proper width to bind into the major groove of the DNA double helix. In helix-turn-helix and helix-loop-helix motifs, an α helix called the recognition helix makes contact with and recognizes a base sequence along the major groove of the DNA (Figure 17.3a and b). Recall that the major groove is a region of the DNA double helix where the bases contact the surrounding water in the cell. Hydrogen bonding between the amino acid side chains in an α helix and nucleotide bases in the DNA is one way that a transcription factor can bind to a specific DNA sequence. In addition, the recognition helix often contains many positively charged amino acids (e.g., arginine and lysine) that favorably interact with the DNA backbone, which is negatively charged. Such basic domains are a common feature of many DNA-binding proteins.

A zinc finger motif is composed of one α-helix and two β-sheet structures that are held together by a zinc (Zn^{2+}) metal



(a) Helix–turn–helix motif

(b) Helix–loop–helix motif

(c) Zinc finger motif

(d) Leucine zipper motif

FIGURE 17.3 **Structural motifs found in transcription factor proteins.** Certain types of protein secondary structure are found in many different transcription factors. In this figure, α helices are shown as cylinders and β sheets as flattened arrows. **(a)** Helix-turn-helix motif: Two α helices are connected by a turn. The α helices lie in the major groove of the DNA. **(b)** Helix-loop-helix motif: A short α helix is connected to a longer α helix by a loop. In this illustration, a dimer is formed from the interactions of two helix-loop-helix motifs, and the longer helices are binding to the DNA. **(c)** Zinc finger motif: Each zinc finger is composed of one α helix and two antiparallel β sheets. A zinc ion (Zn^{2+}), shown in red, holds the zinc finger together. This illustration shows four zinc fingers in a row. **(d)** Leucine zipper motif: The leucine zipper promotes the dimerization of two transcription factor proteins. Two α helices (termed a *coiled coil*) are intertwined via the leucines (see inset).

Concept check: *Explain how an α helix is able to function as a recognition helix in a transcription factor protein.*

ion (Figure 17.3c). The zinc finger also can recognize DNA sequences within the major groove.

A second interesting feature of certain motifs is that they promote protein dimerization. The leucine zipper (Figure 17.3d) and helix-loop-helix motif (see Figure 17.3b) mediate protein dimerization. For example, Figure 17.3d depicts the dimerization and DNA binding of two proteins that have several leucine amino acids (a zipper). Alternating leucines in both proteins interact ("zip up"), resulting in protein dimerization. Two identical transcription factor proteins come together to form a **homodimer,** or two different transcription factors can form a **heterodimer.** As discussed later, the dimerization of transcription factors can be an important way to modulate their function.

Regulatory Transcription Factors Recognize Regulatory Elements That Function as Enhancers or Silencers

As mentioned previously, when the binding of a regulatory transcription factor to a regulatory element increases transcription, the regulatory element is known as an enhancer. Such elements can stimulate transcription 10- to 1000-fold, a phenomenon known as **up regulation.** Alternatively, regulatory elements that serve to inhibit transcription are called silencers, and their action is called **down regulation.**

Many regulatory elements are **orientation-independent,** or **bidirectional.** This means that the regulatory element can function in the forward or reverse orientation. For example, let's consider an enhancer with a forward orientation as shown below:

5′–GATA–3′
3′–CTAT–5′

This enhancer is also bound by a regulatory transcription factor and enhances transcription even when it is oriented in the reverse orientation:

5′–TATC–3′
3′–ATAG–5′

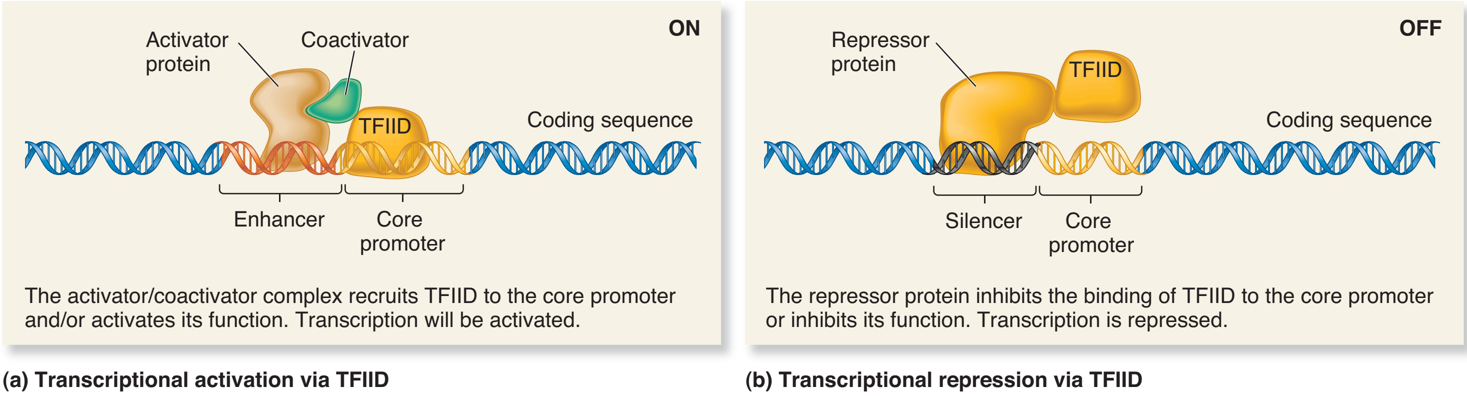
Striking variation is also found in the location of regulatory elements relative to a gene's promoter. Regulatory elements are often located in a region within a few hundred base pairs (bp) upstream from the promoter site. However, they can be quite distant from the promoter, even 100,000 bp away, yet exert strong effects on the ability of RNA polymerase to initiate transcription at the core promoter! Regulatory elements were first discovered by Susumu Tonegawa and coworkers in the 1980s. While studying genes that play a role in immunity, they identified a region that is far away from the core promoter, but is needed for high levels of transcription to take place. In some cases, regulatory elements are located downstream from the promoter site and may even be found within introns, the noncoding parts of genes. As you may imagine, the variation in regulatory element orientation and location profoundly complicates the efforts of geneticists to identify the regulatory elements that affect the expression of any given gene.

Regulatory Transcription Factors May Exert Their Effects Through TFIID and Mediator

Different mechanisms have been discovered that explain how a regulatory transcription factor can bind to a regulatory element, thereby affecting gene transcription. For most genes, more than one mechanism is involved. The net effect of a regulatory transcription factor is to influence the ability of RNA polymerase to transcribe a given gene. However, most regulatory transcription factors do not bind directly to RNA polymerase. How then do these regulatory transcription factors exert their effects? For structural genes in eukaryotes, regulatory transcription factors commonly influence the function of RNA polymerase II by interacting with other proteins that directly control RNA polymerase II. Two protein complexes that communicate the effects of regulatory transcription factors are transcription factor IID (TFIID) and mediator.

As discussed in Chapter 14, **TFIID** is a general transcription factor that binds to the TATA box and is needed to recruit RNA polymerase II to the core promoter. Some regulatory transcription factors bind to a regulatory element and then influence the function of TFIID. Activator proteins are expected to enhance the ability of TFIID to initiate transcription. One possibility is that activator proteins could help recruit TFIID to the TATA box or they could enhance the function of TFIID in a way that facilitates its ability to recruit RNA polymerase II. In some cases, activator proteins exert their effects by interacting with **coactivators**—proteins that increase the rate of transcription but do not directly bind to the DNA itself. This type of activation is shown in **Figure 17.4a.** In contrast, repressors inhibit the function of TFIID. They could exert their effects by preventing the binding of TFIID to the TATA box (**Figure 17.4b**) or by inhibiting the ability of TFIID to recruit RNA polymerase II to the core promoter.

A second way that regulatory transcription factors control RNA polymerase II is via mediator—a protein complex



The activator/coactivator complex recruits TFIID to the core promoter and/or activates its function. Transcription will be activated.

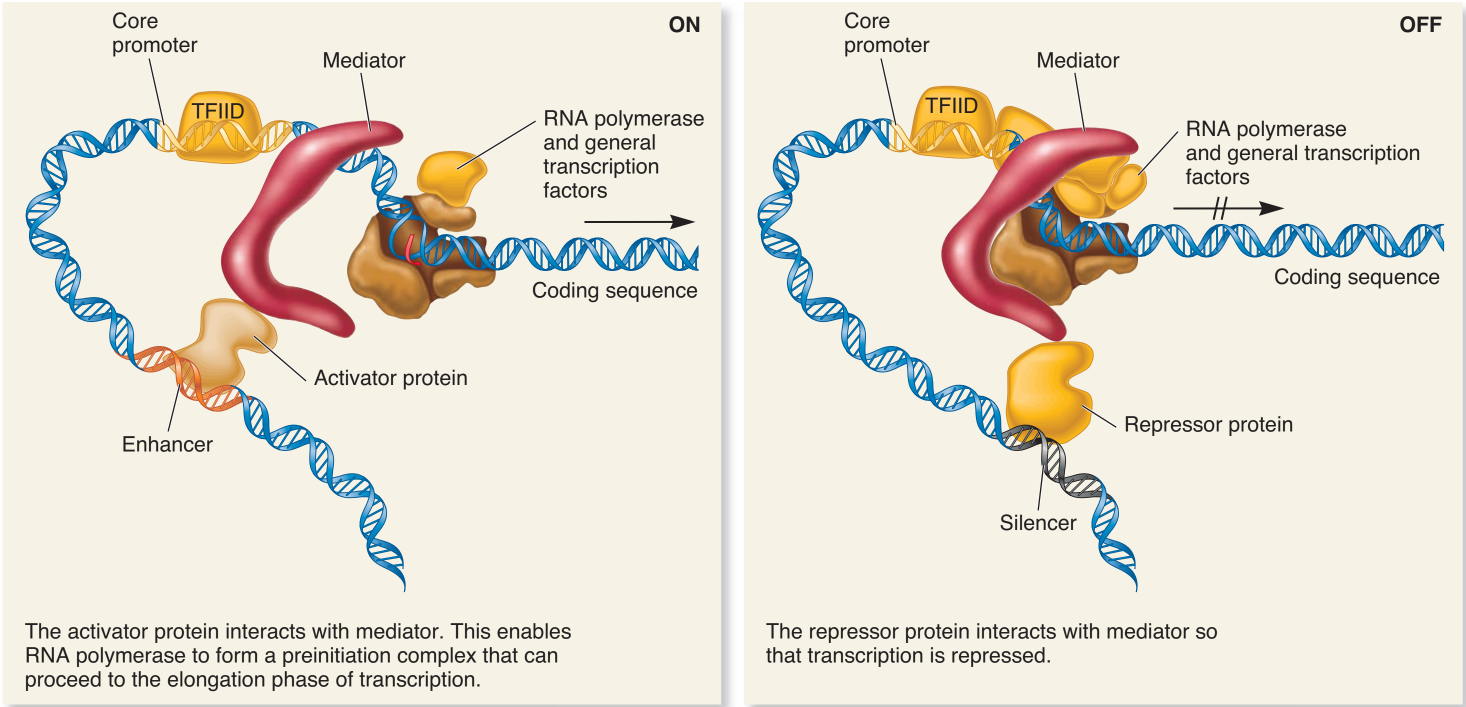
The repressor protein inhibits the binding of TFIID to the core promoter or inhibits its function. Transcription is repressed.

(a) Transcriptional activation via TFIID

(b) Transcriptional repression via TFIID

FIGURE 17.4 **Effects of regulatory transcription factors on TFIID.** **(a)** Some activators stimulate the function of TFIID, thereby enhancing transcription. In this example, the activator interacts with a coactivator that directly binds to TFIID and stimulates its function. **(b)** Regulatory transcription factors may also function as repressors. In the example shown here, the repressor binds to a silencer and inhibits the ability of TFIID to bind to the TATA box at the core promoter.

Concept check: *If a repressor prevents TFIID from binding to the TATA box, why does this inhibit transcription?*


Core promoter
ON
TFIID
Mediator
RNA polymerase and general transcription factors
Coding sequence
Activator protein
Enhancer
The activator protein interacts with mediator. This enables RNA polymerase to form a preinitiation complex that can proceed to the elongation phase of transcription.
Core promoter
OFF
TFIID
Mediator
RNA polymerase and general transcription factors
Coding sequence
Repressor protein
Silencer
The repressor protein interacts with mediator so that transcription is repressed.

(a) Transcriptional activation via mediator

(b) Transcriptional repression via mediator

FIGURE 17.5 **Effects of regulatory transcription factors on mediator.** Some activators interact with mediator in a way that causes RNA polymerase to proceed to the elongation phase of transcription. **(a)** In this example, the enhancer is relatively far away from the core promoter. Therefore, a loop must form in the DNA so the activator can interact with mediator. **(b)** Alternatively, repressors can interact with mediator, thereby preventing RNA polymerase from proceeding to the elongation phase of transcription.

Concept check: *When an activator protein interacts with mediator, how does this affect the function of RNA polymerase?*

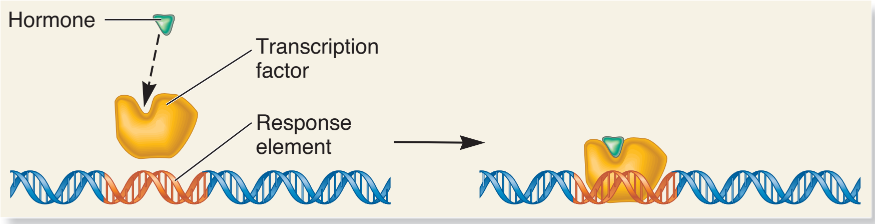
discovered by Roger Kornberg and colleagues in 1990. The term **mediator** refers to the observation that it mediates the interaction between RNA polymerase II and regulatory transcription factors. As discussed in Chapter 14, mediator controls the ability of RNA polymerase II to progress to the elongation stage of transcription. Transcriptional activators stimulate the ability of mediator to facilitate the switch between the initiation and elongation stages, whereas repressors have the opposite effect. In this example shown in **Figure 17.5**, an activator binds to a distant enhancer element. The activator protein and mediator are brought together by the formation of a loop within the intervening DNA.

A third way that regulatory transcription factors can influence transcription is by recruiting proteins to the promoter region that affect nucleosome positions and compositions. For example, certain transcriptional activators recruit proteins that facilitate the conversion of chromatin from a closed to an open conformation. We will return to this topic later in this chapter.

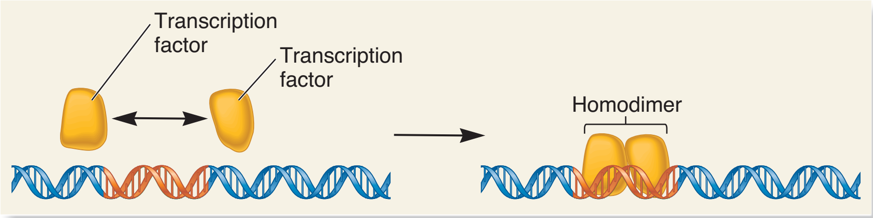
The Function of Regulatory Transcription Factor Proteins Can Be Modulated in Three Ways

Thus far, we have considered the structures of regulatory transcription factors and the molecular mechanisms that account for their abilities to control transcription. The functions of the regulatory transcription factors themselves must also be modulated. Why is this necessary? The answer is that the genes they control must be turned on at the proper time, in the correct cell type, and under the appropriate environmental conditions. Eukaryotes have evolved different ways to modulate the functions of these proteins.

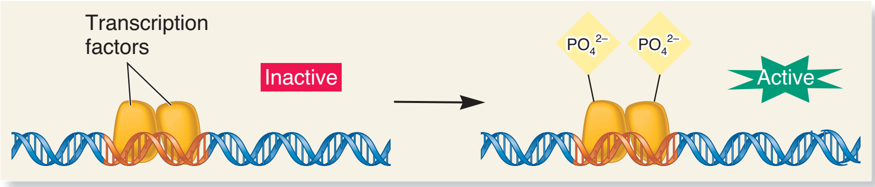
The functions of regulatory transcription factor proteins are controlled in three common ways: through (1) the binding of small effector molecules, (2) protein-protein interactions, and (3) covalent modifications. **Figure 17.6** depicts these three mechanisms of modulating regulatory transcription factor function. Usually, one or more of these modulating effects are important in determining whether a transcription factor can bind to the DNA or influence transcription by RNA polymerase. For example, a small effector molecule may bind to a regulatory transcription factor and promote its binding to DNA (Figure 17.6a). We will see that steroid hormones function in this manner. Another important way is via protein-protein interactions (Figure 17.6b). The formation of homodimers and heterodimers is a fairly common means of controlling transcription. Finally, the function of a regulatory transcription factor can be affected by covalent modifications such as the attachment of a phosphate group (Figure 17.6c).



(a) Binding of a small effector molecule such as a hormone



(b) Protein–protein interaction



(c) Covalent modification such as phosphorylation

FIGURE 17.6 Common ways to modulate the function of regulatory transcription factors. **(a)** The binding of an effector molecule such as a hormone may influence the ability of a transcription factor to bind to the DNA. **(b)** Protein-protein interactions among transcription factor proteins may influence their functions. **(c)** Covalent modifications such as phosphorylation may alter transcription factor function.

Steroid Hormones Exert Their Effects by Binding to a Regulatory Transcription Factor

Now that we have a general understanding of the structure and function of transcription factors, let's turn our attention to a specific example that illustrates how a regulatory transcription factor carries out its role within living cells. Our example is a regulatory transcription factor that responds to steroid hormones.

The ultimate action of a steroid hormone is to affect gene transcription. In animals, steroid hormones act as signaling molecules that are synthesized by endocrine glands and secreted into the bloodstream. The hormones are then taken up by cells that can respond to the hormones in different ways. For example, glucocorticoid hormones influence nutrient metabolism in most body cells. Other steroid hormones, such as estrogen and testosterone, promote the development of secondary sexual characteristics and play a role in reproduction.

Figure 17.7 shows the stepwise action of glucocorticoid hormones, which are produced in mammals. In this example, the hormone enters the cytosol of a cell by diffusing through the plasma membrane. Once inside, the hormone specifically binds to **glucocorticoid receptors.** Prior to hormone binding, the glucocorticoid receptor is complexed with proteins known as heat shock proteins (HSP), one example being HSP90. After the hormone binds to the glucocorticoid receptor, HSP90 is released. This exposes a nuclear localization signal (NLS) that directs a protein into the nucleus. Two glucocorticoid receptors form a homodimer and then travel through a nuclear pore into the nucleus.

How does the glucocorticoid receptor regulate the expression of particular genes? In the nucleus, the glucocorticoid receptor homodimer binds to a glucocorticoid response element (GRE) with two copies of the following consensus sequence:

5′–AGRACA–3′
3′–TCYTGT–5′

where R is a purine and Y is pyrimidine. A GRE is found next to many genes and functions as an enhancer. The binding of the glucocorticoid receptor homodimer to a GRE activates the transcription of the nearby gene, eventually leading to the synthesis of the encoded protein.

Mammalian cells usually have a large number of glucocorticoid receptors within the cytoplasm. Because GREs are located near dozens of different genes, the uptake of many hormone molecules can activate many glucocorticoid receptors, thereby stimulating the transcription of many different genes. For this reason, a cell can respond to the presence of the hormone in a very complex way. Glucocorticoid hormones stimulate many genes that encode proteins involved in several different cellular processes, including the synthesis of glucose, the breakdown of proteins, and the mobilization of fats. Although the genes are not physically adjacent to each other, the regulation of multiple genes via glucocorticoid hormones is much like the ability of bacterial operons to simultaneously control the expression of several genes.

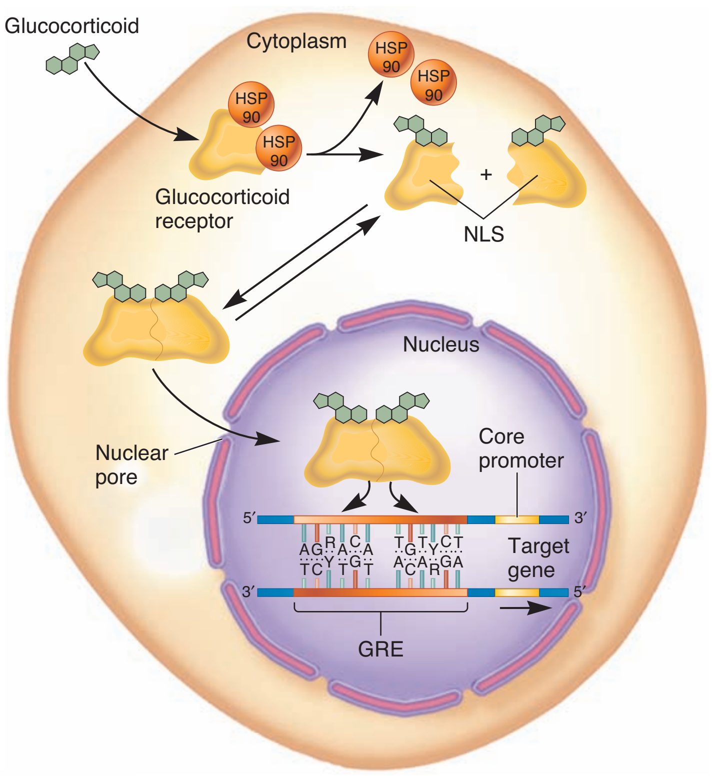


FIGURE 17.7 **The action of glucocorticoid hormones.** Once inside the cell, the glucocorticoid hormone binds to the glucocorticoid receptor, releasing it from a protein known as HSP90. This exposes a nuclear localization signal (NLS). Two glucocorticoid receptors then form a dimer and travel into the nucleus, where the dimer binds to a glucocorticoid response element (GRE) that is next to a particular gene. The binding of the dimer to the GRE activates the transcription of the adjacent target gene.

Genes→Traits Glucocorticoid hormones are produced by the endocrine glands in response to fasting and activity. They enable the body to regulate its metabolism properly. When glucocorticoids are produced, they are taken into cells and bind to the glucocorticoid receptors. This eventually leads to the activation of genes that encode proteins involved in the synthesis of glucose, the breakdown of proteins, and the mobilization of fats.

Concept check: Explain why the glucocorticoid receptor binds specifically next to the core promoter of certain genes, but not next to the core promoter of most genes.

17.1 REVIEWING THE KEY CONCEPTS

- Regulatory transcription factors can be activators that bind to enhancers and increase transcription, or they can be repressors that bind to silencers and inhibit transcription (see Figure 17.2).
- Combinatorial control means that the transcription of a gene is controlled by a variety of factors.
- Transcription factor proteins contain specific domains that may be involved in a variety of processes such as DNA binding and protein dimerization (see Figure 17.3).
- Control or response elements are usually orientation-independent.
- Regulatory transcription factors may exert their effects by interacting with TFIID (a general transcription factor) or mediator (see Figures 17.4, 17.5).
- The functions of regulatory transcription factors can be modulated by small effector molecules, protein-protein interactions, and covalent modifications (see Figure 17.6).
- Steroid hormones, such as glucocorticoids, bind to receptors that function as transcriptional activators (see Figure 17.7).

17.1 COMPREHENSION QUESTIONS

1. Combinatorial control refers to the phenomenon that
   a. transcription factors always combine with each other when regulating genes.
   b. the combination of many factors determines the expression of any given gene.
   c. small effector molecules and regulatory transcription factors are found in many different combinations.
   d. genes and regulatory transcription factors must combine with each other during gene regulation.
2. A regulatory transcription factor protein typically contains _________ that binds to the _________ of the DNA.
   a. an α helix, backbone
   b. an α helix, major groove
   c. a β sheet, backbone
   d. a β sheet, major groove
3. A bidirectional enhancer has the following sequence:

   5′–GTCA–3′
   3′–CAGT–5′

   Which of the following sequences would also be a functional enhancer?

   a. 5′–ACTG–3′
      3′–TGAC–5′
   b. 5′–TGAC–3′
      3′–ACTG–5′
   c. 3′–GTCA–5′
      5′–CAGT–3′
   d. 3′–TGAC–5′
      5′–ACTG–3′
4. Regulatory transcription factors can be modulated by
   a. the binding of small effector molecules.
   b. protein-protein interactions.
   c. covalent modifications.
   d. all of the above.

17.2 CHROMATIN REMODELING, HISTONE VARIATION, AND HISTONE MODIFICATION

In Chapter 12, we considered how eukaryotic DNA is complexed with proteins, such as histones, to form chromatin. The term **ATP-dependent chromatin remodeling,** or simply **chromatin remodeling,** refers to dynamic changes in the structure of chromatin that occur during the life of a cell. These changes range from local alterations in the positioning of one or a few nucleosomes to larger changes that affect chromatin structure over a longer distance. Chromatin remodeling is carried out by ATP-dependent chromatin-remodeling complexes, which are a set of diverse multiprotein machines that reposition and restructure nucleosomes.

In eukaryotes, changes in nucleosome position and histone composition are key features of gene regulation. The regulation of transcription depends not only on the activity of regulatory transcription factors that influence RNA polymerase via TFIID and mediator to turn genes on or off, but must also involve changes in chromatin structure that affect the ability of transcription factors to gain access to and bind their target sequences in the promoter region. If the chromatin is in a **closed conformation,** transcription may be difficult or impossible. By comparison, chromatin that is in an **open conformation** is more easily accessible to transcription factors and RNA polymerase so that transcription can occur. Although the closed and open conformations may be affected by the relative compaction of a chromosomal region, researchers have recently determined that histone composition and the precise positioning of nucleosomes at or near promoters often play a key role in eukaryotic gene regulation. In this section, we examine the molecular mechanisms that explain how changes in chromatin structure controls the regulation of eukaryotic genes.

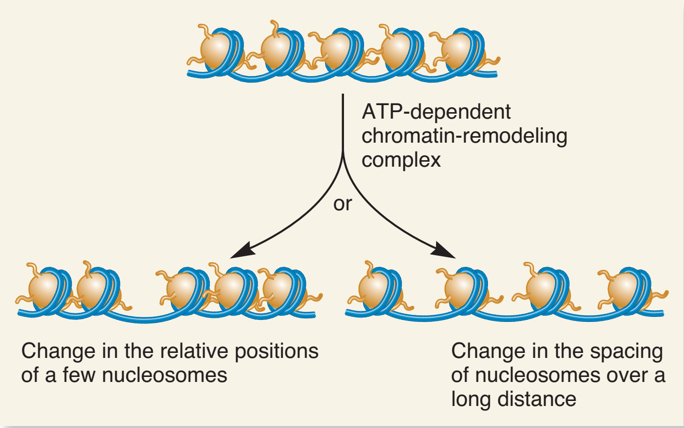
Chromatin Remodeling Complexes Alter the Positions and Compositions of Nucleosomes

In recent years, geneticists have been trying to identify the steps that promote the interconversion between the closed and open conformations of chromatin. Nucleosomes have been shown to change position in cells that normally express a particular gene compared with cells in which the gene is inactive. For example, in red blood cells that express the β-globin gene, an alteration in nucleosome positioning occurs in the promoter region from nucleotide −500 to +200. This alteration is thought to be an important step so the β-globin gene can be expressed. Based on the analysis of many genes, researchers have discovered that a key role of some transcriptional activators is to orchestrate changes in chromatin structure from the closed to the open conformation by altering nucleosomes.

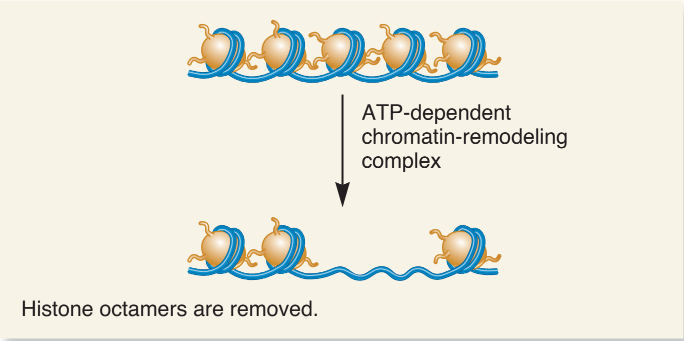
One way to change chromatin structure is through ATP-dependent chromatin remodeling. In this process, the energy derived from ATP hydrolysis is used to drive a change in the locations and/or compositions of nucleosomes, thereby making the DNA more or less amenable to transcription. Therefore, chromatin remodeling is important for both the activation and repression of transcription.

The remodeling process is carried out by an ATP-dependent chromatin-remodeling complex, a protein complex that recognizes nucleosomes and uses ATP to alter their configuration. All remodeling complexes have motor proteins called **DNA translocases** that move along the DNA. Eukaryotes have multiple families of chromatin remodelers. Though their names may differ depending on the species, common ATP-dependent chromatin-remodeling complex families are referred to as the SWI/SNF family, the ISWI family, the INO80 family, and the Mi-2 family. The names of these remodelers sometimes refer to the effects of mutations in genes that encode remodeling proteins. For example, the abbreviations SWI and SNF refer to the effects that occur in yeast when these remodeling enzymes are defective. SWI mutants are defective in mating-type switching, and SNF mutations create a sucrose nonfermenting phenotype.

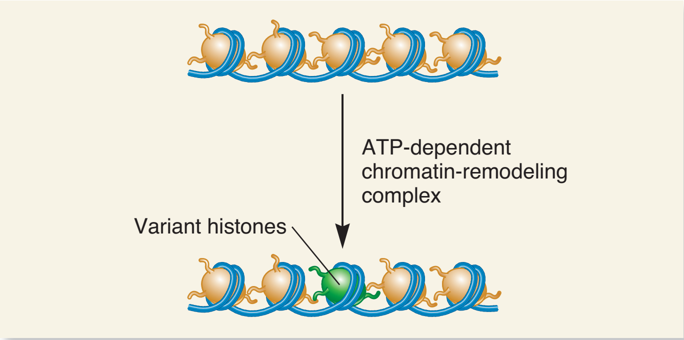
How do chromatin remodelers change chromatin structure? Three effects are possible. One result of ATP-dependent chromatin remodeling is a change in the location of nucleosomes (**Figure 17.8a**). This may involve a shift in nucleosomes to a new location or a change in the relative spacing of nucleosomes over a


ATP-dependent chromatin-remodeling complex
or
Change in the relative positions of a few nucleosomes
Change in the spacing of nucleosomes over a long distance

(a) Change in nucleosome position


ATP-dependent chromatin-remodeling complex
Histone octamers are removed.

(b) Histone eviction



(c) Replacement with variant histones

FIGURE 17.8 **ATP-dependent chromatin remodeling.** ATP-dependent chromatin-remodeling complexes may **(a)** change the locations of nucleosomes, **(b)** remove histones from the DNA, or **(c)** replace core histones with variant histones.

Concept check *How might nucleosome eviction affect transcription?*

long stretch of DNA. A second effect is that remodelers may evict histones from the DNA, thereby creating gaps where nucleosomes are not found (**Figure 17.8b**). A third possibility is that remodelers may change the composition of nucleosomes by removing standard histones and replacing them with histone variants (**Figure 17.8c**). The functions of histone variants are described next.

Histone Variants Play Specialized Roles in Chromatin Structure and Function

As discussed in Chapter 12, the five histone genes, which encode histones H1, H2A, H2B, H3, and H4, are moderately repetitive. The total number of histone genes varies from species to species. For example, the human genome contains over 70 histone genes that have been produced by gene duplication events during evolution. Most of the histone genes encode standard histone proteins. However, a few have accumulated mutations that change the amino acid sequence of the histone proteins. These are called **histone variants.** Among eukaryotic species, histone variants have been identified for histone H1, H2A, H2B, and H3 genes, but not for the H4 gene.

What are the consequences of histone variation? Research over the past two decades has shown that certain histone variants play specialized roles in chromatin structure and function. In all eukaryotes, histone variants are incorporated into a subset of nucleosomes to create functionally specialized regions of chromatin. In most cases, the standard histones are incorporated into the nucleosomes while new DNA is synthesized during S phase of the cell cycle. Later, some of the standard histones are replaced by histone variants via chromatin-remodeling complexes.

Table 17.1 describes the standard histones and a few histone variants that are found in humans. A key role of many histone variants is to regulate the structure of chromatin, thereby influencing gene transcription. Such variants can have opposite effects. The incorporation of histone H2A.Bbd into a chromosomal region where a particular gene is found favors gene activation. In contrast, the incorporation of histone $H1^0$ represses gene expression.

Although our focus in this chapter is on gene regulation, it is worth noting that histone variants play other important roles. For example, histone cenH3 (also called CENP-A), which is a variant of histone H3, is found at the centromere of each chromosome and functions in the binding of kinetochore proteins. Histone cenH3 is required for the proper segregation of eukaryotic chromosomes. Other histone variants are primarily found at specialized sites in certain cells. Histone macroH2A is found along the inactivated X chromosome in female mammals, whereas spH2B is found at the telomeres in sperm cells. Finally, certain histone variants appear to play a role in DNA repair. For example, histone H2A.X becomes phosphorylated where a double-stranded DNA break occurs. This phosphorylation is thought to be important for the proper repair of that break.

TABLE 17.1

Standard Human Histones and Examples of Histone Variants

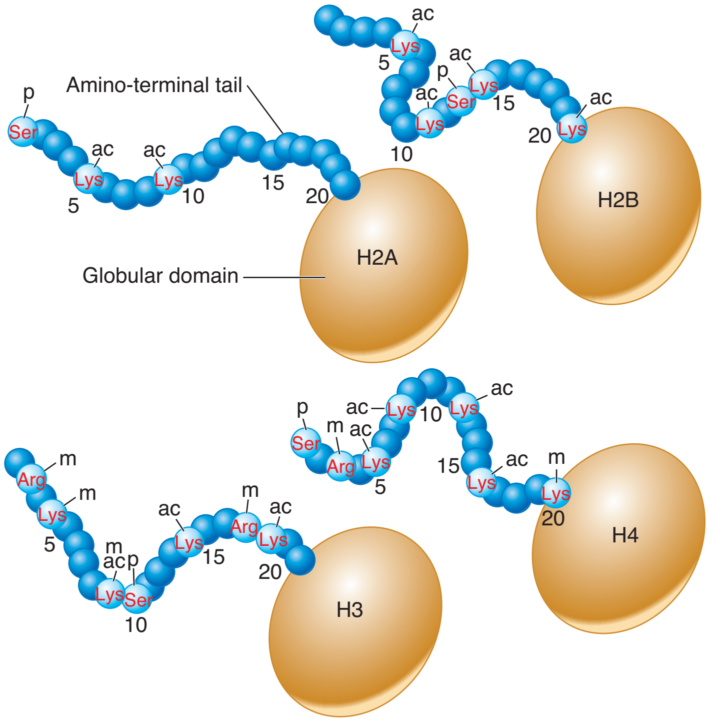
Histone	Type	Number of Genes in Humans	Function
H1	Standard	11	Standard linker histone*
$H1^0$	Variant	1	Linker histone associated with chromatin compaction and gene repression
H2A	Standard	15	Standard core histone
MacroH2A	Variant	1	Core histone that is abundant on the inactivated X chromosome in female mammals. Plays a role in chromatin compaction
H2A.Z	Variant	1	Core histone that is usually found in nucleosomes that flank the transcriptional start site of promoters. Plays a role in gene transcription
H2A.Bbd	Variant	1	Core histone that promotes open chromatin. Plays a role in gene activation
H2A.X	Variant	1	Plays a role in DNA repair
H2B	Standard	17	Standard core histone
spH2B	Variant	1	Core histone found in the telomeres of sperm cells
H3	Standard	10	Standard core histone
cenH3	Variant	1	Core histone found at centromeres. Involved with the binding of kinetochore proteins
H3.3	Variant	2	Core histone that promotes open chromatin. Plays a role in gene activation
H4	Standard	14	Standard core histone

*H1 in mammals is found in five subtypes.

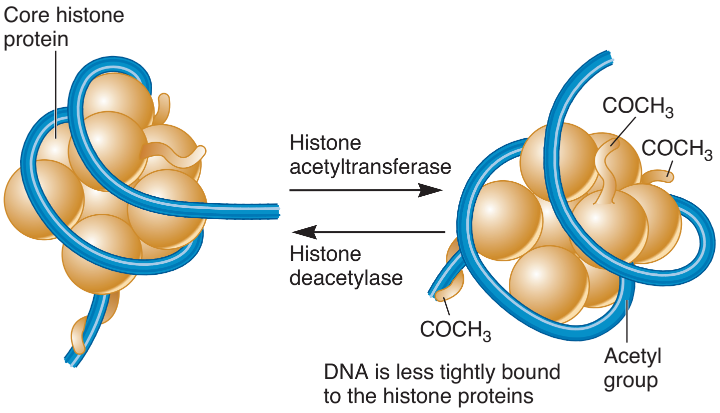
The Histone Code Also Controls Gene Transcription

As described in Chapter 12, each of the core histone proteins consists of a globular domain and a flexible, charged amino terminus called an amino-terminal tail (refer back to Figure 12.10a). The DNA wraps around the globular domains, and the amino-terminal tails protrude from the chromatin. In recent years, researchers have discovered that particular amino acids in the amino-terminal tails of standard histones and histone variants are subject to several types of covalent modifications, including acetylation, methylation, and phosphorylation. Over 50 different histone-modifying enzymes have been identified in mammals that selectively modify amino-terminal tails of histones. **Figure 17.9a** shows examples of sites in the tails of H2A, H2B, H3, and H4 that can be modified.

How do histone modifications affect the level of transcription? First, they may directly influence interactions between histones and the DNA. For example, positively charged lysines within the core histone proteins can be acetylated by enzymes called **histone acetyltransferases.** The attachment of the acetyl group (—$COCH_3$) eliminates the positive charge on the lysine side chain, thereby disrupting the electrostatic attraction between the histone protein and the negatively charged DNA backbone (**Figure 17.9b**).



(a) Examples of histone modifications



(b) Effect of acetylation

FIGURE 17.9 Histone modifications and their effects on nucleosome structure. **(a)** Examples of histone modifications that may occur at the amino-terminal tail of the four core histone proteins. The abbreviations refer to p (phosphate), ac (acetyl), and m (methyl) groups. **(b)** Effect of acetylation. When the core histones are acetylated via histone acetyltransferase, the DNA becomes less tightly bound to the histones. Histone deacetylase removes the acetyl groups.

Concept check: Describe two different ways that histone modifications may alter chromatin structure.

In addition, histone modifications occur in patterns that are recognized by proteins. According to the **histone code hypothesis,** proposed by Brian Strahl, C. David Allis, and Bryan Turner in 2000, the pattern of histone modification acts much like a language or code in specifying alterations in chromatin structure. For example, one pattern might involve phosphorylation of the serine at the first position in H2A and acetylation of the fifth and eighth amino acids in H4, which are lysines. A different pattern could involve acetylation of the fifth amino acid, a lysine, in H2B and methylation of the third amino acid in H4, which is an arginine.

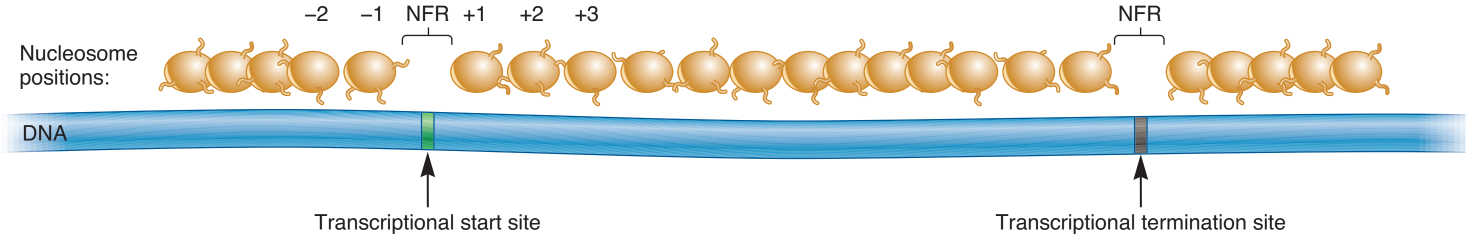
The pattern of covalent modifications to the amino-terminal tails provides binding sites for proteins that subsequently affect the degree of transcription. One pattern of histone modification may attract proteins that inhibit transcription, which would silence the transcription of genes in the region. A different combination of histone modifications may attract proteins, such as chromatin-remodeling enzymes, which would serve to alter the positions of nucleosomes in a way that promotes gene transcription. For example, the acetylation of histones attracts certain chromatin remodelers that can shift or evict nucleosomes, thereby aiding in the transcription of genes. Overall, the histone code is thought to play an important role in determining whether the information within the genomes of eukaryotic species is accessed. Researchers are trying to decipher the effects of the covalent modifications that make up the histone code.

Eukaryotic Genes Are Flanked by Nucleosome-Free Regions and Well-Positioned Nucleosomes

Studies over the last 10 years or so have revealed that many eukaryotic genes show a common pattern of nucleosome organization (Figure 17.10). For active genes or those genes that can be activated, the core promoter is found at a **nucleosome-free region (NFR),** which is a site that is missing nucleosomes. The NFR is typically 150 bp in length. Although the NFR may be required for transcription, it is not, by itself, sufficient for gene activation. At any given time in the life of a cell, many genes that contain an NFR are not being actively transcribed.

FIGURE 17.10 Nucleosome arrangements and composition in the vicinity of a structural gene.

Concept check: Why is an NFR needed at the core promoter for transcription to occur?



A nucleosome-free region (NFR) is found at the beginning and end of many genes. Nucleosomes tend to be precisely positioned near the beginning and end of a gene, but are less regularly distributed elsewhere.

The NFR is flanked by two well-positioned nucleosomes that are termed the −1 and +1 nucleosomes. In yeast, the transcriptional start site (TSS) is usually at the boundary between the NFR and the +1 nucleosome. However, in animals, the TSS is about 60 bp farther upstream into the NFR. The +1 nucleosome typically contains histone variants H2A.Z and H3.3. Depending on the species and the gene, these variants may also be found in the −1 nucleosome and in some of the nucleosomes that immediately follow the +1 nucleosome in the transcribed region. For example, the +2 nucleosome is likely to contain H2A.Z but not as likely as the +1 nucleosome. Similarly, the +3 nucleosome is likely to contain H2A.Z, but not as likely as the +2 nucleosome.

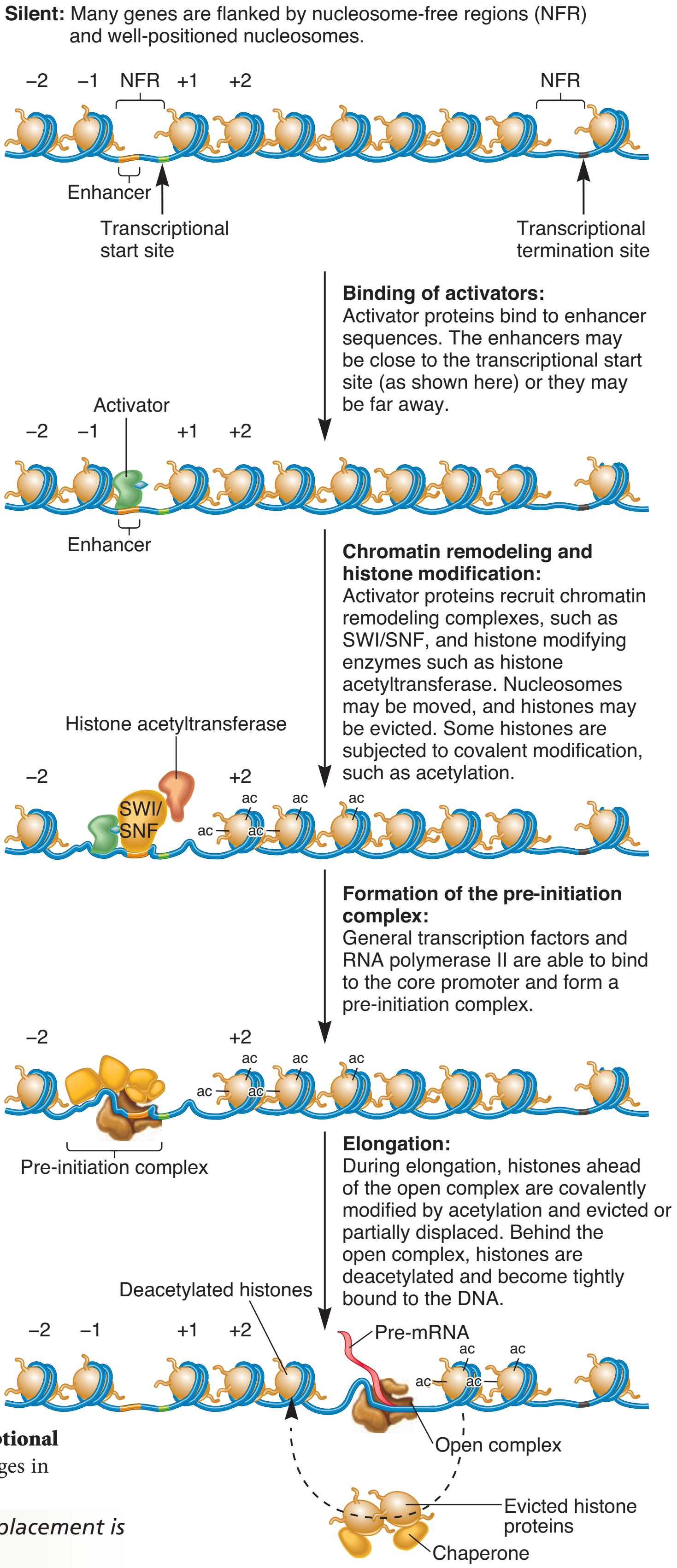
The nucleosomes downstream from the +1 nucleosome tend to be evenly spaced near the beginning of a eukaryotic gene, but their spacing becomes less regular farther downstream. The end of many eukaryotic genes has a well-positioned nucleosome that is followed by an NFR. This arrangement at the end of genes may be important for transcriptional termination.

Transcriptional Activation Involves Changes in Nucleosome Locations, Composition, and Histone Modifications

A key role of certain transcriptional activators is to recruit chromatin-remodeling enzymes and histone-modifying enzymes to the promoter region. Though the order of recruitment may differ among specific transcriptional activators, this appears to be critical for transcriptional initiation and elongation. **Figure 17.11** presents a general scheme for how transcriptional activators may facilitate transcription in a eukaryotic gene, such as a gene found in yeast.

As discussed earlier, a gene that is able to be activated has an NFR at the transcriptional start site. In the scenario shown in Figure 17.11, a transcriptional activator binds to an enhancer in the NFR. The activator then recruits chromatin-remodeling complexes and histone-modifying enzymes to this region. The chromatin remodelers may shift nucleosomes or temporarily evict nucleosomes from the promoter region. Nucleosomes containing the histone variant H2A.Z, which are typically found at the +1 nucleosome, are thought to be more easily removed from the DNA than those containing the standard histone H2A. Histone-modifying enzymes, such as histone acetyltransferase, covalently modify histone proteins and may affect nucleosome contact with the DNA. The actions of chromatin remodelers and histone-modifying enzymes facilitate the binding of general transcription factors and RNA polymerase II to the core promoter, thereby allowing the formation of a preinitiation complex (see Figure 17.11).

Further changes in chromatin structure are necessary for elongation to occur. RNA polymerase II cannot transcribe DNA



Silent: Many genes are flanked by nucleosome-free regions (NFR) and well-positioned nucleosomes.

Binding of activators: Activator proteins bind to enhancer sequences. The enhancers may be close to the transcriptional start site (as shown here) or they may be far away.

Chromatin remodeling and histone modification: Activator proteins recruit chromatin remodeling complexes, such as SWI/SNF, and histone modifying enzymes such as histone acetyltransferase. Nucleosomes may be moved, and histones may be evicted. Some histones are subjected to covalent modification, such as acetylation.

Formation of the pre-initiation complex: General transcription factors and RNA polymerase II are able to bind to the core promoter and form a pre-initiation complex.

Elongation: During elongation, histones ahead of the open complex are covalently modified by acetylation and evicted or partially displaced. Behind the open complex, histones are deacetylated and become tightly bound to the DNA.

ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 17.11 **A simplified model for the transcriptional activation of a eukaryotic gene.** The activation involves changes in nucleosome arrangements and histone modifications.

Concept check: Explain why histone eviction or displacement is needed for the elongation phase of transcription.

that is tightly wrapped in nucleosomes. For transcription to occur, histones are evicted, partially displaced, or destabilized so that RNA polymerase II can pass. Evicted histones are transferred to histone chaperones, which are proteins that bind histones and aid in the assembly of histones. Assembled histones are then placed back on the DNA behind the moving RNA polymerase II (see Figure 17.11).

Histone-modifying enzymes also play a key role in histone removal and replacement during the elongation phase of transcription. Histone-modifying enzymes have been found to travel with RNA polymerase II during the elongation phase of transcription. These include enzymes that carry out histone acetylation, H3 methylation, and H2B ubiquitination. These modifications facilitate histone removal ahead of the traveling RNA polymerase II. Behind RNA polymerase II, histone deacetylase removes the acetyl groups, thereby favoring the binding of histones to the DNA to form nucleosomes. The re-formation of nucleosomes behind RNA polymerase II is thought to be critical to maintaining the fidelity of transcription. If nucleosome re-formation did not occur properly, transcriptional initiation might occur at multiple points in a gene, thereby producing faulty transcripts.

Changes in the relative amounts of histone variants also occur within actively transcribed genes. For example, histone variant H3.3 is often found in the transcribed regions of genes, but is less common in silent genes. H3.3 may facilitate the eviction and replacement of nucleosomes ahead and behind RNA polymerase II, respectively. Also, genes with very high levels of transcription may be largely devoid of histones because multiple RNA polymerases are transcribing them.

17.2 REVIEWING THE KEY CONCEPTS

- Chromatin remodeling occurs via ATP-dependent chromatin-remodeling complexes that alter the position and composition of nucleosomes (see Figure 17.8).
- Histone variants, which have amino acid sequences that differ slightly from the standard histones, play specialized roles in chromatin structure and function (see Table 17.1).
- Amino-terminal histone tails are subject to covalent modifications that act as a histone code for the binding of proteins that affect chromatin structure and gene expression (see Figure 17.9).
- Many eukaryotic genes are flanked by a nucleosome-free region (NFR) with well-positioned nucleosomes (see Figure 17.10).
- Transcriptional activation involves changes in nucleosome locations, nucleosome composition, and histone modifications (see Figure 17.11).

17.2 COMPREHENSION QUESTIONS

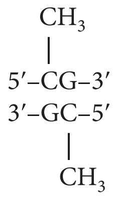
1. A chromatin-remodeling complex may
   a. change the locations of nucleosomes.
   b. evict nucleosomes from DNA.
   c. replace standard histones with histone variants.
   d. do all of the above.
2. According to the histone code hypothesis, the pattern of histone modifications acts like a language that
   a. influences chromatin structure.
   b. promotes transcriptional termination.
   c. inhibits the elongation of RNA polymerase.
   d. does all of the above.
3. Which of the following characteristics is typical of a eukaryotic gene that can be transcribed?
   a. The core promoter is wrapped around a nucleosome.
   b. The core promoter is found in a nucleosome-free region.
   c. The terminator is wrapped around a nucleosome.
   d. None of the above are correct.
4. Transcriptional activation of eukaryotic genes involves which of the following events?
   a. Changes in nucleosome locations
   b. Changes in histone composition within nucleosomes
   c. Changes in histone modifications.
   d. All of the above

17.3 DNA METHYLATION

We now turn our attention to a regulatory mechanism that usually silences gene expression. As discussed in previous chapters, DNA structure can be modified by the covalent attachment of methyl groups, a mechanism called **DNA methylation.** This process is common in some, but not all, eukaryotic species. For example, yeast and *Drosophila* have little or no detectable methylation of their DNA, whereas DNA methylation in vertebrates and plants is relatively abundant. In mammals, approximately 2% to 7% of the DNA is methylated. In this section, we will examine how DNA methylation occurs and how it can control gene expression.

DNA Methylation Occurs on the Cytosine Base and Usually Inhibits Gene Transcription

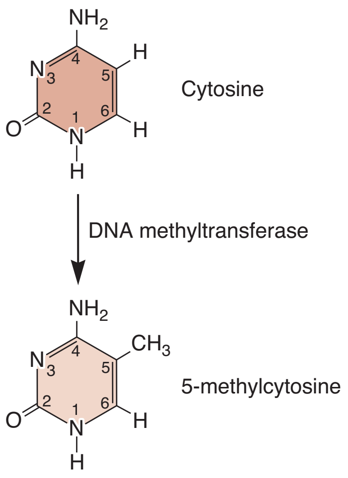
As shown in **Figure 17.12**, eukaryotic DNA methylation occurs via an enzyme called **DNA methyltransferase,** which attaches a methyl group to the number 5 position of the cytosine base, forming 5-methylcytosine. The sequence that is methylated is shown here.



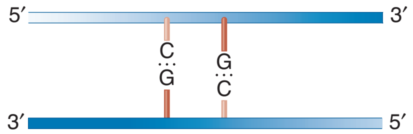
$$\begin{array}{l} \;\;\;\;\;\mathrm{CH_3} \\ \;\;\;\;\;\;| \\ 5'\text{–CG–}3' \\ 3'\text{–GC–}5' \\ \;\;\;\;\;\;\;\;\;\;| \\ \;\;\;\;\;\;\;\;\mathrm{CH_3} \end{array}$$

Note that this sequence contains cytosines in both strands. Methylation of the cytosine in both strands is termed full methylation, whereas methylation of only one strand is called hemimethylation. The methyl group on cytosine protrudes into the major groove of the DNA and may affect the binding of proteins into the major groove.

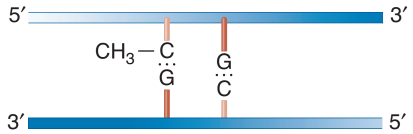
DNA methylation usually inhibits the transcription of eukaryotic genes, particularly when it occurs in the vicinity of the promoter. In vertebrates and plants, **CpG islands** occur near



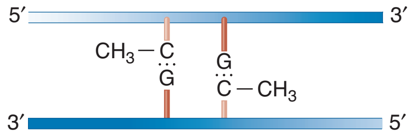
(a) The methylation of cytosine



(b) Unmethylated



(c) Hemimethylated



(d) Fully methylated

FIGURE 17.12 **DNA methylation on cytosine bases.**
(a) Methylation occurs via an enzyme known as DNA methyltransferase, which attaches a methyl group to the number 5 carbon on cytosine. The CG sequence can be **(b)** unmethylated, **(c)** hemimethylated, or **(d)** fully methylated.

many promoters of genes. (Note: CpG refers to a dinucleotide of C and G in DNA that is connected by a phosphodiester linkage.) These CpG islands are commonly 1000 to 2000 bp in length and contain a high number of CpG sites. In the case of **housekeeping genes**—genes that encode proteins required in most cells of a multicellular organism—the cytosine bases in the CpG islands are unmethylated. Therefore, housekeeping genes tend to be expressed in most cell types. By comparison, other genes are highly regulated and may be expressed only in a particular cell type. These are termed **tissue-specific genes.** In some cases, it has been found that the expression of tissue-specific genes may be silenced by the methylation of CpG islands. In this way, DNA methylation is thought to play an important role in the silencing of tissue-specific genes to prevent them from being expressed in the wrong tissue.

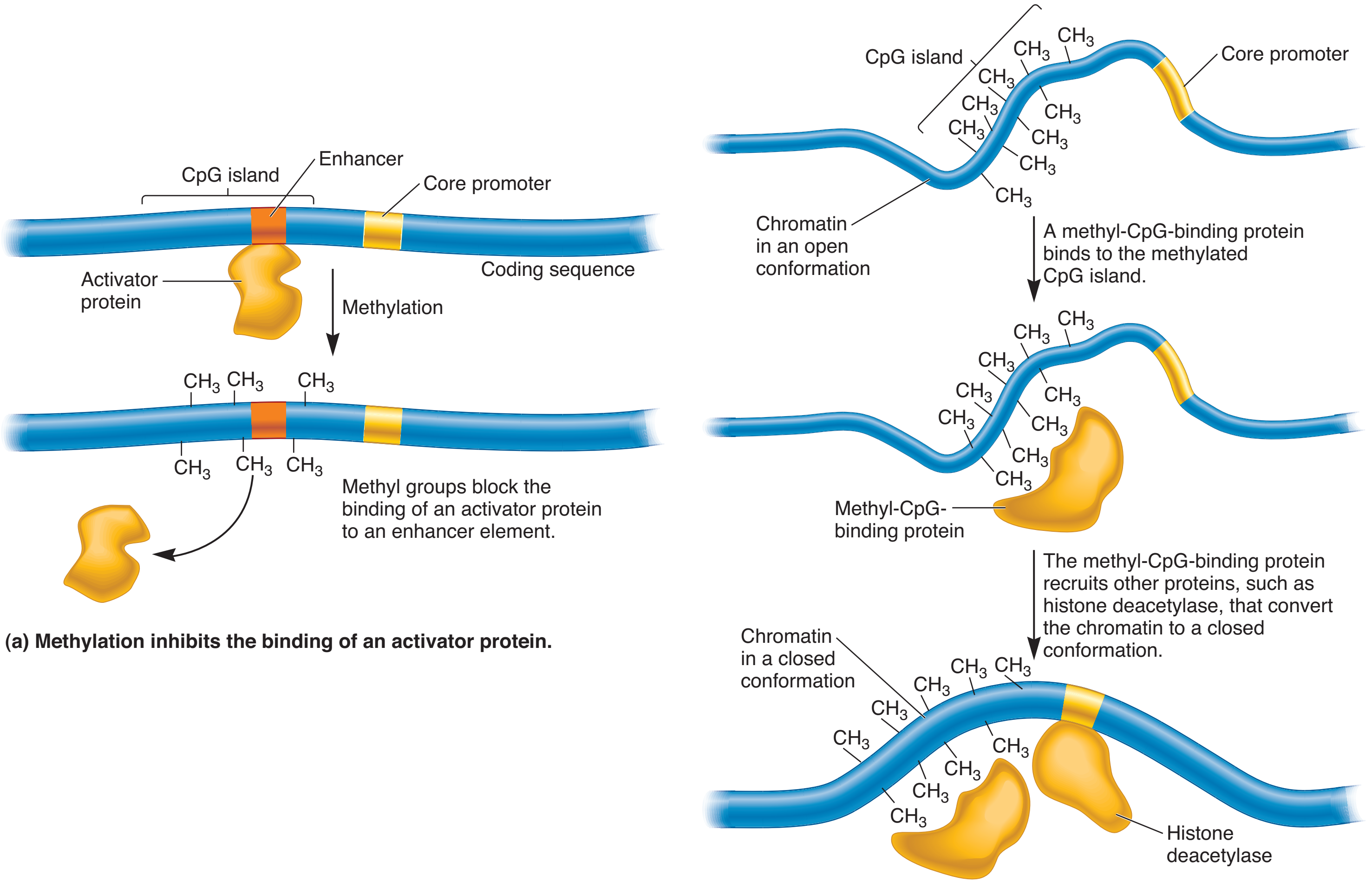
Methylation can affect transcription in two general ways. First, methylation of CpG islands may prevent or enhance the binding of regulatory transcription factors to the promoter region. For example, methylated CG sequences could prevent the binding of an activator protein to an enhancer element, presumably by the methyl group protruding into the major groove of the DNA (Figure 17.13a). The inability of an activator protein to bind to the DNA inhibits the initiation of transcription. However, CG methylation does not slow down the movement of RNA polymerase along a gene. In vertebrates and plants, coding regions downstream from the core promoter usually contain methylated CG sequences, but these do not hinder the elongation phase of transcription. This suggests that methylation must occur in the vicinity of the promoter to have an effect on transcription.

A second way that methylation inhibits transcription is via proteins known as **methyl-CpG-binding proteins,** which bind methylated sequences (Figure 17.13b). These proteins contain a domain called the methyl-binding domain that specifically recognizes a methylated CG sequence. Once bound to the DNA, the methyl-CpG-binding proteins recruit other proteins to the region that inhibit transcription. For example, methyl-CpG-binding proteins may recruit histone deacetylase to a methylated CpG island near a promoter. Histone deacetylation removes acetyl groups from the histone proteins, which makes it more difficult for nucleosomes to be removed from the DNA. In this way, deacetylation tends to inhibit transcription.

DNA Methylation Is Heritable

Methylated DNA sequences are inherited during cell division. Experimentally, if fully methylated DNA is introduced into a plant or vertebrate cell, the DNA will remain fully methylated even in subsequently produced daughter cells. However, if the same sequence of nonmethylated DNA is introduced into a cell, it will remain nonmethylated in the daughter cells. These observations indicate that the pattern of methylation is retained following DNA replication and, therefore, is inherited in future daughter cells.

How can methylation be inherited from cell to cell? Figure 17.14 illustrates a molecular model that explains this process, which was originally proposed by Arthur Riggs, Robin Holliday, and J. E. Pugh in 1975. The DNA in a particular cell may become methylated by **de novo methylation**—the methylation of DNA that was previously unmethylated. When a fully methylated segment of DNA replicates in preparation for cell division, the newly made daughter strands contain unmethylated cytosines. Because only one strand is methylated, such DNA is said to be hemimethylated. This hemimethylated DNA is recognized by DNA methyltransferase, which makes it fully methylated. This process is called **maintenance methylation,** because it preserves the methylated condition in future cells. However, maintenance methylation does not act on unmethylated DNA. Overall, maintenance methylation appears to be an efficient process that routinely occurs within vertebrate and plant cells. By comparison, de novo methylation and


CpG island
Enhancer
Core promoter
Activator protein
Coding sequence
Methylation
CH_3 CH_3 CH_3
CH_3 CH_3 CH_3
Methyl groups block the binding of an activator protein to an enhancer element.
CpG island
CH_3
Core promoter
Chromatin in an open conformation
A methyl-CpG-binding protein binds to the methylated CpG island.
Methyl-CpG-binding protein
The methyl-CpG-binding protein recruits other proteins, such as histone deacetylase, that convert the chromatin to a closed conformation.
Chromatin in a closed conformation
Histone deacetylase

(a) Methylation inhibits the binding of an activator protein.

(b) Methyl-CpG-binding protein recruits other proteins that change the chromatin to a closed conformation.

FIGURE 17.13 **Transcriptional silencing via methylation.** **(a)** The methylation of a CpG island may inhibit the binding of transcriptional activators to the promoter region. **(b)** The binding of a methyl-CpG-binding protein to a CpG island may lead to the recruitment of proteins, such as histone deacetylase, that convert chromatin to a closed conformation and thus suppress transcription.

Concept check: *Explain why the events shown in part (a) inhibit transcription.*

demethylation are infrequent and highly regulated events. According to this view, the initial methylation or demethylation of a given gene can be regulated so it occurs in a specific cell type or stage of development. Once methylation has occurred, it can then be transmitted from mother to daughter cells via maintenance methylation.

The methylation mechanism shown in Figure 17.14 can explain a form of epigenetic inheritance called genomic imprinting, which is described in Chapter 6. In this case, specific genes are methylated during oogenesis or spermatogenesis, but not both. Following fertilization, the pattern of methylation is maintained in the offspring. For example, if a gene is methylated only during spermatogenesis, the allele that is inherited from the father will be methylated in the somatic cells of the offspring, but the maternal allele will remain unmethylated. Along these lines, geneticists are also eager to determine how variations in DNA methylation patterns may be important for cell differentiation. It may be a key way to silence genes in different cell types. However, additional research is necessary to understand how specific genes may be targeted for de novo methylation or demethylation during different developmental stages or in specific cell types.

17.3 REVIEWING THE KEY CONCEPTS

- DNA methylation is the attachment of a methyl group to a cytosine base via DNA methyltransferase (see Figure 17.12).
- The methylation of CpG islands near promoters usually silences transcription. Methylation may affect the binding of regulatory transcription factors or it may inhibit transcription via methyl-CpG-binding proteins (see Figure 17.13).
- Maintenance methylation can preserve a methylation pattern in daughter cells following cell division (see Figure 17.14).

17.3 COMPREHENSION QUESTIONS

1. How can methylation affect transcription?
   a. It may prevent the binding of regulatory transcription factors.

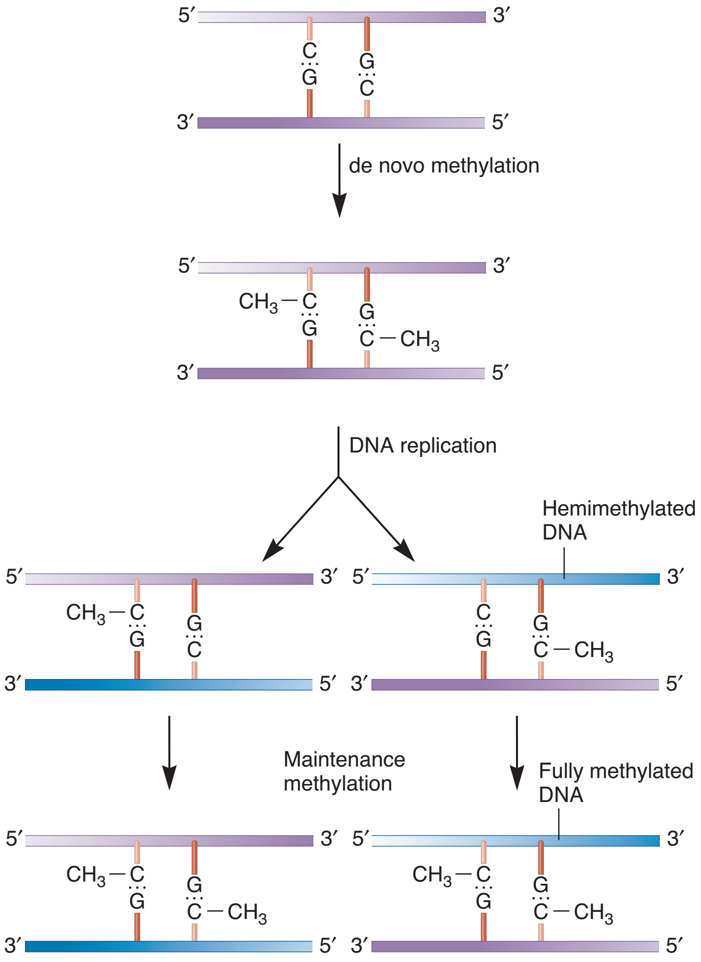


FIGURE 17.14 **A molecular model for the inheritance of DNA methylation.** The DNA initially undergoes de novo methylation, which is a rare, highly regulated event. Once this occurs, DNA replication produces hemimethylated DNA molecules, which are then fully methylated by DNA methyltransferase. This process, called maintenance methylation, is a routine event that is expected to occur for all hemimethylated DNA.

Concept check: What is the difference between de novo methylation and maintenance methylation?

b. It may enhance the binding of regulatory transcription factors.
c. It may attract the binding of methyl-CpG-binding proteins, which inhibit transcription.
d. All of the above are possible.

2. The process in which completely unmethylated DNA becomes methylated is called
   a. maintenance methylation.
   b. de novo methylation.
   c. primary methylation.
   d. demethylation.

17.4 INSULATORS

Thus far, we have considered how regulatory transcription factors, chromatin remodeling, and DNA methylation can regulate gene transcription. As we have seen, eukaryotic gene regulation involves changes in chromatin structure that may occur over a relatively long distance. In addition, gene regulation often involves regulatory elements such as enhancers that are far away from the promoters they regulate.

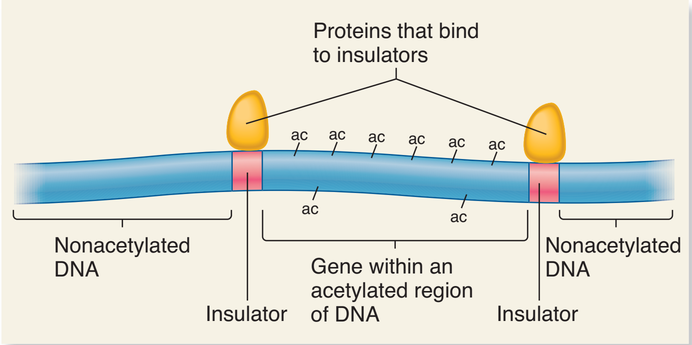
In eukaryotes, adjacent genes usually exhibit different patterns of gene regulation. Because eukaryotic gene regulation can occur over long distances, a bewildering aspect of such regulation is its ability to control a particular gene but not affect neighboring genes. How are the effects of chromatin remodeling and histone modifications constrained so they affect only a particular gene and not its neighbors? How is an enhancer prevented from controlling multiple genes?

An **insulator** is a segment of DNA that functions as a boundary between two genes. An insulator is so named because it protects, or insulates, a gene from the regulatory effects of a neighboring gene. In this section, we will consider how insulators are able to carry out this critical function.

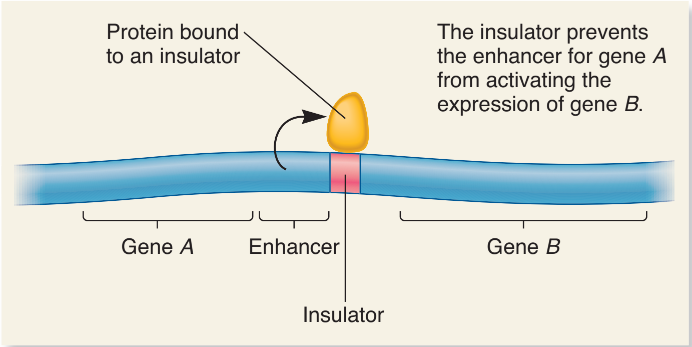
Insulators May Act as a Barrier to Changes in Chromatin Structure or Block the Effects of Neighboring Enhancers

Insulators typically perform two roles. Some insulators act as a barrier to chromatin-remodeling or histone-modifying enzymes. As an example, let's consider the effects of histone deacetylase, which removes acetyl groups from core histones, thereby favoring a closed chromatin conformation that is transcriptionally silent. Histone deacetylase may act over a long region of chromatin. How is the action of histone deacetylase controlled so that certain genes in a chromosomal region are not silenced? In the example of Figure 17.15a, a gene is found in a chromosomal region in which most of the histones are not acetylated due to the action of histone deacetylase. However, this gene is flanked by two insulators that allow the region where the gene is located to be highly acetylated and transcriptionally activated. In this example, the insulators act as barriers to the action of histone deacetylase. The mechanisms that enable insulators to act as barriers are not well understood. However, such insulators often bind proteins that recruit histone-modifying enzymes or chromatin-remodeling enzymes to the region. For example, an insulator could bind a protein that recruits histone acetyltransferase, which would favor the acetylation of core histones.

A second role of insulators is to block the effects of enhancers that exert their effects on neighboring genes. As discussed earlier in this chapter, enhancers may stimulate the expression of genes that are relatively far away. To prevent an enhancer of one gene from activating the expression of an adjacent gene, an insulator may be located between them. In the example shown in Figure 17.15b, the enhancer can activate the expression of gene *A* but the insulator prevents it from exerting its effects on gene *B*. In other words, gene *B* is insulated from the effects of this enhancer. How does this occur? Although the mechanism of enhancer blocking may vary among different genes, one mechanism is chromosome looping, which is described next.


Proteins that bind to insulators
ac
ac
ac
ac
ac
ac
ac
ac
Nonacetylated DNA
Gene within an acetylated region of DNA
Nonacetylated DNA
Insulator
Insulator

(a) Insulators as a barrier to changes in chromatin structure


Protein bound to an insulator
The insulator prevents the enhancer for gene *A* from activating the expression of gene *B*.
Gene *A*
Enhancer
Gene *B*
Insulator

(b) Insulator that blocks the effects of a neighboring enhancer

FIGURE 17.15 **Potential effects of insulators.** **(a)** Some insulators act as a barrier to enzymes that move along the DNA and change chromatin structure or composition. The label ac refers to acetyl groups attached to histone proteins. **(b)** Other insulators prevent enhancers of adjacent genes from exerting their effects.

Concept check: Why are insulators important for gene regulation in eukaryotes?

Insulators May Promote the Formation of Loops That Block the Effects of Nearby Enhancers

How do insulators block the regulatory effects of enhancers? The mechanisms of insulator function are still being actively investigated, but one way they may block the activity of enhancers is by a looping mechanism. This type of mechanism is thought to operate with regard to an insulator that affects the *Igf2* gene, which is discussed in Chapter 6 (see Figure 6.11). The *Igf2* gene is paternally imprinted. Therefore, only the gene copy from the father is active; the allele inherited from the mother is not. **Figure 17.16** shows a simplified mechanism for the effects of an insulator (also called the imprinting control region, or ICR) that is located between the *H19* and *Igf2* genes. An enhancer is found next to the *H19* gene. To prevent this enhancer from also stimulating the *Igf2* gene, a loop is formed (see Figure 17.16 top right). For this to occur, a protein called the CTC-binding factor binds to both the insulator and to a site that follows the *Igf2* gene called a differentially methylated region (DMR). In the maternally inherited *Igf2* gene, the CTC-binding factors bind to these sites and then bind to each other to form a loop in the DNA. This loop blocks the ability of the enhancer to stimulate the *Igf2* gene, rendering it transcriptionally silent.

Alternatively, if the ICR and DMR are methylated, which occurs during sperm formation, CTC-binding factor is unable to bind to the DNA (see bottom right of Figure 17.16). This prevents loop formation, which allows the enhancer to stimulate the *Igf2* gene. Therefore, the paternally inherited *Igf2* gene is transcriptionally activated.

17.4 REVIEWING THE KEY CONCEPTS

- An insulator is a segment of DNA that functions as a boundary that protects one gene from the regulatory effects of a neighboring gene. Insulators may act as a barrier to changes

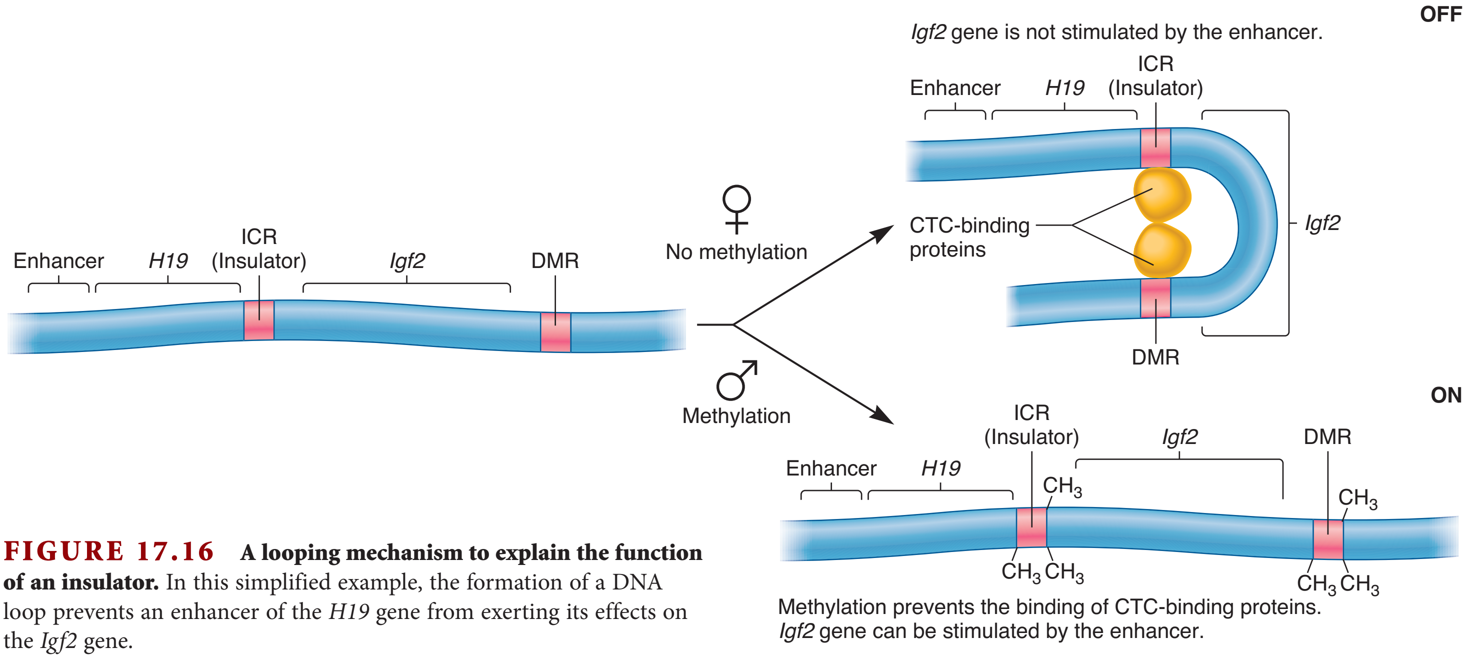

OFF
Igf2 gene is not stimulated by the enhancer.
Enhancer
H19
ICR (Insulator)
Igf2
CTC-binding proteins
DMR
No methylation
Methylation
Enhancer
H19
ICR (Insulator)
Igf2
DMR
ON
Enhancer
H19
ICR (Insulator)
Igf2
DMR
CH_3
CH_3 CH_3
CH_3
CH_3 CH_3
Methylation prevents the binding of CTC-binding proteins. *Igf2* gene can be stimulated by the enhancer.

FIGURE 17.16 **A looping mechanism to explain the function of an insulator.** In this simplified example, the formation of a DNA loop prevents an enhancer of the *H19* gene from exerting its effects on the *Igf2* gene.

in chromatin structure or block the effects of neighboring enhancers (see Figure 17.15).

- Insulators may block the effects of enhancers by promoting the formation of a loop in the DNA (see Figure 17.16).

17.4 COMPREHENSION QUESTIONS

1. Insulators may exert their effect by
   a. acting as a barrier to changes in chromatin structure.
   b. blocking the effects of neighboring enhancers.
   c. both a and b.
   d. none of the above.
2. An imprinting control region (ICR) is a type of insulator that binds a protein called ______________that binds to ___________ DNA.
   a. methyl-CpG binding protein, unmethylated
   b. methyl-CpG binding protein, methylated
   c. CTC-binding factor, unmethylated
   d. CTC-binding factor, methylated

17.5 REGULATION OF RNA PROCESSING, RNA STABILITY, AND TRANSLATION

Thus far, we have considered a variety of mechanisms that regulate the level of gene transcription. These mechanisms control the amount of RNA transcribed from a given gene. In addition, eukaryotic gene expression is commonly regulated after the RNA is made (**Table 17.2**). One mechanism is pre-mRNA processing. Following transcription, a pre-mRNA transcript is processed before it becomes a functional mRNA. These processing events, which include splicing, capping, polyA tailing, and RNA editing, were described in Chapter 14. We begin this section by examining how the process of alternative splicing is regulated at the RNA level.

TABLE **17.2**

Gene Regulation via RNA Processing and Translation

Effect	Description
Alternative splicing	Certain pre-mRNAs can be spliced in more than one way, leading to polypeptides that have different amino acid sequences. Alternative splicing is often cell-specific so that a protein can be fine-tuned to function in a particular cell type. It is an important form of gene regulation in multicellular eukaryotic species.
RNA stability	The amount of RNA is greatly influenced by the half-life of RNA transcripts. A long polyA tail on mRNAs promotes their stability due to the binding of polyA-binding protein. Some RNAs with a relatively short half-life contain sequences that target them for rapid degradation. Some RNAs are stabilized by specific RNA-binding proteins that usually bind near the 3′ end.
RNA interference	Double-stranded RNA can mediate the degradation of specific mRNAs in the cell or prevent them from being translated. This is a mechanism of gene regulation. Also, it probably provides eukaryotic cells with protection from invasion by certain types of viruses and may prevent the movement of transposable elements.
Translational regulation of mRNAs	Some mRNAs are regulated via binding proteins that inhibit the ability of the ribosomes to initiate translation. These proteins usually bind at the 5′ end of the mRNA, thereby preventing the ribosome from binding.

Another strategy for regulating gene expression is to influence the concentration of mRNA. As discussed earlier in this chapter, this can be accomplished by regulating the rate of transcription. When the transcription of a gene is increased, a higher concentration of the corresponding RNA results. In this section, we will consider how RNA concentration is greatly affected by the stability or half-life of a particular RNA. Factors that increase RNA stability are expected to raise the concentration of that RNA molecule. We will explore how sequences within mRNA molecules greatly affect their stability.

We will also consider a mechanism of mRNA silencing known as RNA interference that involves double-stranded RNAs that may direct the breakdown of specific RNAs or inhibit their translation. Last, we will see how the process of mRNA translation may be regulated by RNA-binding proteins that prevent ribosomes from initiating the translation process.

Alternative Splicing Regulates Which Exons Occur in an RNA Transcript, Allowing Different Polypeptides to Be Made from the Same Structural Gene

When it was first discovered, the phenomenon of splicing seemed like a rather wasteful process. During transcription, energy is used to synthesize intron sequences. Likewise, energy is also used to remove introns via a large spliceosome complex. This observation intrigued many geneticists, because natural selection tends to eliminate wasteful processes. Therefore, instead of simply viewing splicing as a wasteful process, many geneticists expected to find that pre-mRNA splicing has one or more important biological roles. In recent years, one very important biological advantage, termed **alternative splicing,** has become apparent. Alternative splicing refers to the phenomenon that a pre-mRNA can be spliced in more than one way.

What is the advantage of alternative splicing? To understand the biological effects of alternative splicing, remember that the sequence of amino acids within a polypeptide determines the structure and function of a protein. Alternative splicing produces two or more polypeptides with differences in their amino acid sequences, leading to possible changes in their functions. In most cases, the alternative versions of the protein have similar functions, because most of their amino acid sequences are identical to each other. Nevertheless, alternative splicing produces differences in amino acid sequences that provide each polypeptide with its own unique characteristics. Because alternative splicing allows two or more different polypeptide sequences to be derived from a single gene, some geneticists have speculated that an important advantage of this process is that it allows an organism to carry fewer genes in its genome.

The degree of splicing and alternative splicing varies greatly among different species. Baker's yeast (*Saccharomyces cerevisiae*), for example, contains about 6300 genes and approximately 300 (i.e., ~5%) encode pre-mRNAs that are spliced. Of these, only a few have been shown to be alternatively spliced. Therefore, in this unicellular eukaryote, alternative splicing is not a major mechanism for generating protein diversity. In comparison, complex multicellular organisms rely on alternative splicing to a great extent. Humans have approximately 20,000 to 25,000 different genes, and most of these contain one or more introns. Recent estimates suggest that about 70% of all human pre-mRNAs are alternatively spliced. Furthermore, certain pre-mRNAs are alternatively spliced to an extraordinary extent. Some pre-mRNAs can be alternatively spliced to produce dozens of different mRNAs. This provides a much greater potential for human cells to produce protein diversity.

Figure 17.17 considers an example of alternative splicing for a gene that encodes a protein known as α-tropomyosin. This protein functions in the regulation of cell contraction. It is located along the thin filaments found in smooth muscle cells, such as those in the uterus and small intestine, and in striated muscle cells that are found in cardiac and skeletal muscle. α-tropomyosin is also synthesized in many types of nonmuscle cells but in lower amounts. Within a multicellular organism, different types of cells must regulate their contractibility in subtly different ways. One way this may be accomplished is by the production of different forms of α-tropomyosin.

The intron-exon structure of rat α-tropomyosin pre-mRNA and two alternative ways that the pre-mRNA can be spliced are shown in Figure 17.17. The pre-mRNA contains 14 exons, 6 of which are **constitutive exons** (shown in red), which are always found in the mature mRNA from all cell types. Presumably, constitutive exons encode polypeptide segments of the α-tropomyosin protein that are necessary for its general structure and function. By comparison, **alternative exons** (shown in green) are not always found in the mRNA after splicing has occurred. The polypeptide sequences encoded by alternative exons may subtly change the function of α-tropomyosin to meet the needs of the cell type in which it is found. For example, Figure 17.17 shows the predominant splicing products found in smooth muscle cells and striated muscle cells. Exon 2 encodes a segment of the α-tropomyosin protein that alters its function to make it suitable for smooth muscle cells. By comparison, the α-tropomyosin mRNA found in striated muscle cells does not include exon 2. Instead, this mRNA contains exon 3, which is more suitable for that cell type.

Alternative splicing is not a random event. Rather, the specific pattern of splicing is regulated in any given cell. The molecular mechanism for the regulation of alternative splicing involves proteins known as **splicing factors.** Such splicing factors play a key role in the choice of particular splice sites. **SR proteins** are an example of a type of splicing factor. SR proteins contain a domain at their carboxyl-terminal end that is rich in serines (S) and arginines (R) and is involved in protein-protein recognition. They also contain an RNA-binding domain at their amino-terminal end.

As discussed in Chapter 14, components of the spliceosome recognize the 5′ and 3′ splice sites and then remove the intervening intron. The key effect of splicing factors is to modulate the ability of the spliceosome to choose 5′ and 3′ splice sites. This can occur in two ways. Some splicing factors act as repressors that inhibit the ability of the spliceosome to recognize a splice site. In **Figure 17.18a**, a splicing repressor binds to a 3′-splice site and prevents the spliceosome from recognizing the site. Instead, the spliceosome binds to the next available 3′-splice site. The splicing repressor causes exon 2 to be spliced out of the mature

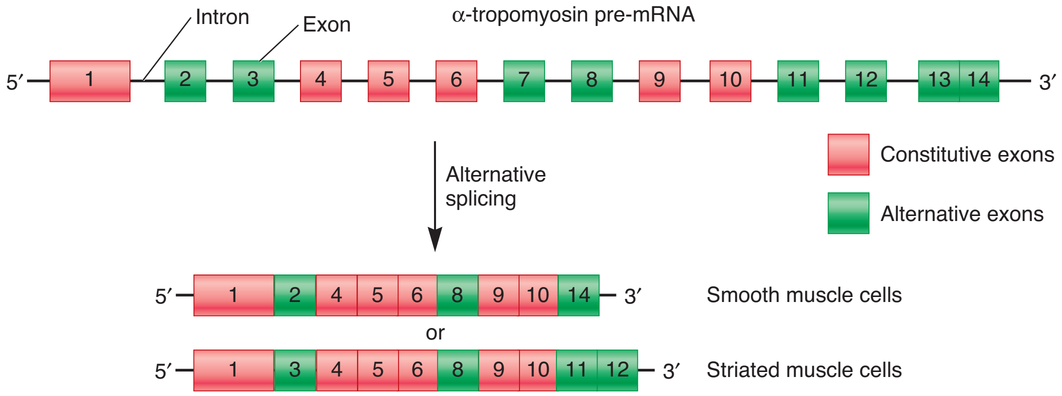
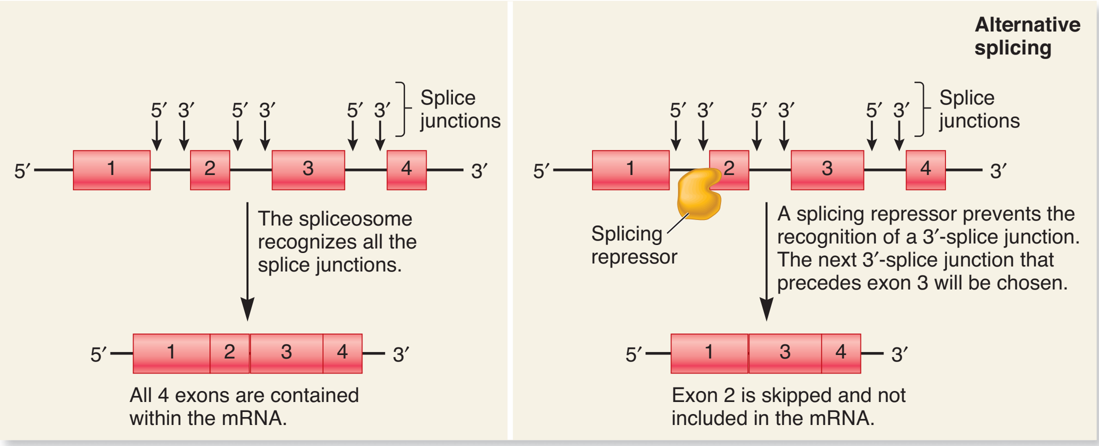
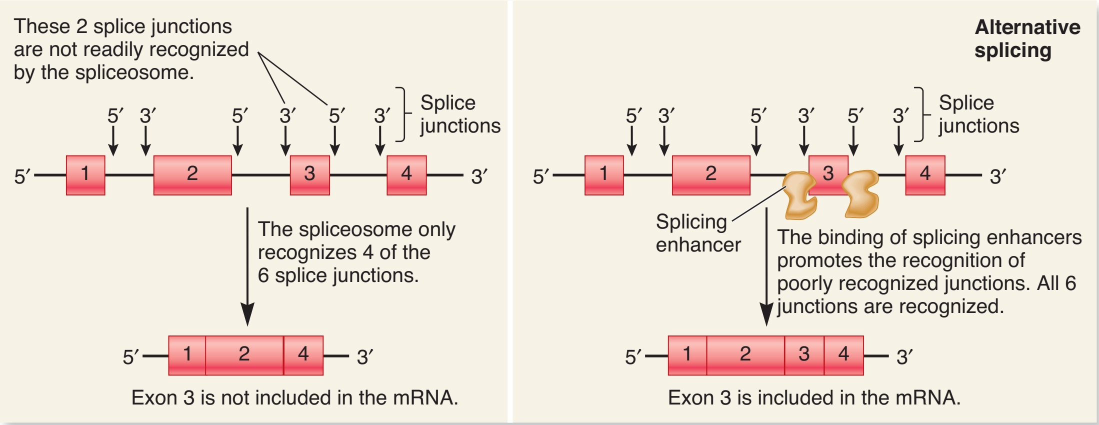


FIGURE 17.17 **Alternative ways that the rat α-tropomyosin pre-mRNA can be spliced.** The top part of this figure depicts the structure of the rat α-tropomyosin pre-mRNA. Exons are shown as colored boxes, and introns are illustrated as connecting black lines. The lower part of the figure describes the final mRNA products in smooth and striated muscle cells. Note: Exon 8 is found in the final mRNA of smooth and striated muscle cells, but not in the mRNA of certain other cell types. The junction between exons 13 and 14 contains a 3′ splice site so that exon 13 can be separated from exon 14.

Genes→Traits α-tropomyosin functions in the regulation of cell contraction in muscle and nonmuscle cells. Alternative splicing of the pre-mRNA provides a way to vary contractibility in different types of cells by modifying the function of α-tropomyosin. As shown here, the alternatively spliced versions of the pre-mRNA produce α-tropomyosin proteins that differ from each other in their structure (i.e., amino acid sequence). These alternatively spliced versions vary in function to meet the needs of the cell type in which they are found. For example, the sequence of exons 1–2–4–5–6–8–9–10–14 produces an α-tropomyosin protein that functions suitably in smooth muscle cells. Overall, alternative splicing affects the traits of an organism by allowing a single gene to encode several versions of a protein, each optimally suited to the cell type in which it is made.


Alternative splicing
5′ 3′ 5′ 3′ 5′ 3′
Splice junctions
5′
1
2
3
4
3′
The spliceosome recognizes all the splice junctions.
5′
1
2
3
4
3′
All 4 exons are contained within the mRNA.
5′ 3′ 5′ 3′ 5′ 3′
Splice junctions
5′
1
2
3
4
3′
Splicing repressor
A splicing repressor prevents the recognition of a 3′-splice junction. The next 3′-splice junction that precedes exon 3 will be chosen.
5′
1
3
4
3′
Exon 2 is skipped and not included in the mRNA.

(a) Splicing repressors


These 2 splice junctions are not readily recognized by the spliceosome.
5′ 3′ 5′ 3′ 5′ 3′
Splice junctions
5′
1
2
3
4
3′
The spliceosome only recognizes 4 of the 6 splice junctions.
5′
1
2
4
3′
Exon 3 is not included in the mRNA.
Alternative splicing
5′ 3′ 5′ 3′ 5′ 3′
Splice junctions
5′
1
2
3
4
3′
Splicing enhancer
The binding of splicing enhancers promotes the recognition of poorly recognized junctions. All 6 junctions are recognized.
5′
1
2
3
4
3′
Exon 3 is included in the mRNA.

(b) Splicing enhancers

FIGURE 17.18 **The roles of splicing factors during alternative splicing.** **(a)** Splicing factors can act as repressors to prevent the recognition of splice sites. In this example, the presence of the splicing repressor causes exon 2 to be skipped and thus not included in the mRNA. **(b)** Other splicing factors can enhance the recognition of splice sites. In this example, the splicing enhancers promote the recognition of sites that flank exon 3, thereby causing its inclusion in the mRNA.

Concept check: *A pre-mRNA is recognized by just one splicing repressor that binds to the 3′-end of the third intron. The third intron is located between exon 3 and exon 4. After splicing is complete, would you expect the mRNA to contain exon 3 and/or exon 4 in the presence of the splicing repressor?*

mRNA, an event called **exon skipping.** Alternatively, other splicing factors enhance the ability of the spliceosome to recognize particular splice sites. In Figure 17.18b, splicing enhancers bind to the 3′- and 5′-splice sites that flank exon 3, which results in the inclusion of exon 3 in the mature mRNA.

Alternative splicing in different tissues is thought to occur because each cell type has its own characteristic concentration of many kinds of splicing factors. Furthermore, much like transcription factors, splicing factors may be regulated by the binding of small effector molecules, protein-protein interactions, and covalent modifications. Overall, the differences in the composition of splicing factors and the regulation of their activities form the basis for alternative splicing decisions.

The Stability of mRNA Influences mRNA Concentration

In eukaryotes, the stability of mRNAs vary considerably. Certain mRNAs have very short half-lives, such as several minutes, whereas others persist for many days or even months. In some cases, the stability of an mRNA is regulated so its half-life is shortened or lengthened. A change in the stability of mRNA can greatly influence the cellular concentration of that mRNA molecule. In this way, mechanisms that control mRNA stability dramatically affect gene expression.

Various factors play a role in mRNA stability. One important structural feature is the length of the polyA tail. As discussed in Chapter 14, most newly made mRNAs contain a polyA tail that averages 200 nucleotides in length. The polyA tail is recognized by the **polyA-binding protein.** The binding of this protein enhances RNA stability. However, as an mRNA ages, its polyA tail tends to be shortened by the action of cellular exonucleases. Once it becomes less than 10 to 30 adenosines in length, the polyA-binding protein can no longer bind, and the mRNA is rapidly degraded by exo- and endonucleases.

Certain mRNAs, particularly those with short half-lives, contain sequences that act as destabilizing elements. Although these destabilizing elements can be located anywhere within the mRNA, they are most commonly located near the 3′ end between the stop codon and the polyA tail. This region of the mRNA is known as the **3′-untranslated region (3′-UTR).** An example of a destabilizing element is the **AU-rich element (ARE)** that is found in many short-lived mRNAs (**Figure 17.19**). This element, which contains the consensus sequence AUUUA, is recognized by cellular proteins that bind to the ARE, thereby influencing whether or not the mRNA is rapidly degraded.

Double-Stranded RNA Can Silence the Expression of mRNA

Another way that specific RNAs can be targeted for degradation or translational inhibition is via a mechanism involving double-stranded RNA. Research in plants and the nematode *Caenorhabditis elegans* led to the discovery that double-stranded RNA can silence the expression of particular genes. Studies of plant viruses were one avenue of research that identified this mechanism of gene regulation. Certain plant viruses produce double-stranded RNA as part of their reproductive cycle. In addition, these plant viruses may carry genes very similar to genes that already exist in the genome of the plant cell. When such viruses infect plant cells, they silence the expression of the plant gene that is similar to the viral gene.

Plant research that involved the production of transgenic plants also supported the idea that double-stranded RNA causes mRNA to be degraded or inhibited. Using molecular techniques described in Chapter 20, cloned genes can be introduced into the genome of plants. Surprisingly, researchers observed that when cloned genes were introduced in multiple copies, the expression of the gene was often silenced. How were these results explained? We now know this may be due to the formation of double-stranded RNA. When a cloned gene randomly inserts into a genome, it may, as a matter of random chance, happen to insert itself next to a promoter for a plant gene that is already present in the genome (**Figure 17.20**). The cloned gene itself has a promoter to make a copy of the sense strand. If it randomly inserts next to a plant gene promoter that is oriented in the opposite direction, both strands of the cloned gene would

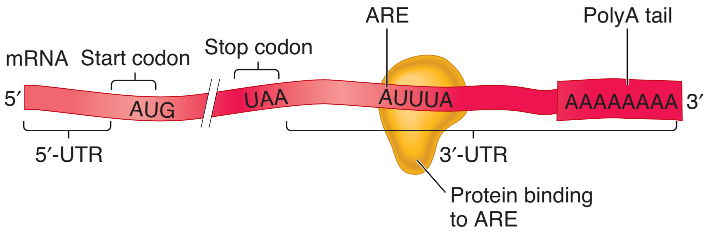

mRNA
Start codon
Stop codon
ARE
PolyA tail
5′
AUG
UAA
AUUUA
AAAAAAAA
3′
5′-UTR
3′-UTR
Protein binding to ARE

FIGURE 17.19 **The location of AU-rich elements (AREs) within mRNAs.** One or more AREs are commonly found within the 3′-UTRs (untranslated regions) of mRNAs with short half-lives. The 5′-UTR is the untranslated region of the mRNA that precedes the start codon. Cellular proteins bind to AREs, thereby influencing whether or not the mRNA is rapidly degraded.

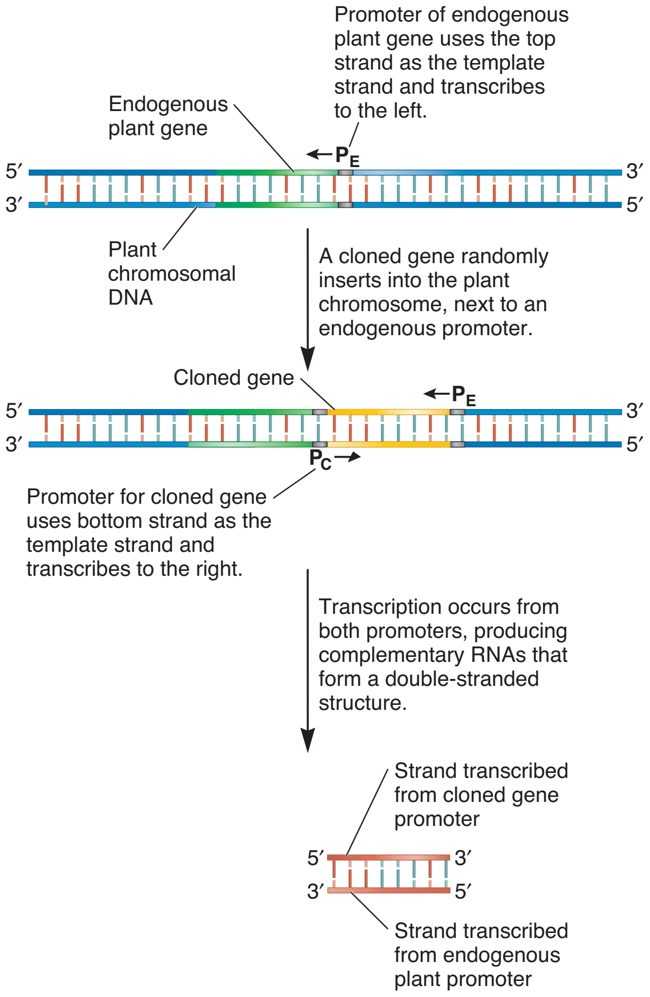

Promoter of endogenous plant gene uses the top strand as the template strand and transcribes to the left.
Endogenous plant gene
P_E
5′
3′
3′
5′
Plant chromosomal DNA
A cloned gene randomly inserts into the plant chromosome, next to an endogenous promoter.
Cloned gene
P_E
5′
3′
3′
5′
P_C
Promoter for cloned gene uses bottom strand as the template strand and transcribes to the right.
Transcription occurs from both promoters, producing complementary RNAs that form a double-stranded structure.
Strand transcribed from cloned gene promoter
5′
3′
3′
5′
Strand transcribed from endogenous plant promoter

FIGURE 17.20 **Gene insertion leading to the production of double-stranded RNA.** When a cloned gene with its own promoter is introduced into a plant cell, it randomly inserts into the chromosomal DNA. In some cases, the cloned gene may insert next to a promoter for an endogenous plant gene, so the promoter for the endogenous gene transcribes the antisense strand of the cloned gene. Therefore, the two promoters transcribe the sense strand and the antisense strand to produce complementary RNAs that will form double-stranded RNA.

Concept check: Why is the production of double-stranded RNA more likely if multiple copies of a gene are inserted into a plant cell's genome?

be transcribed, thereby generating double-stranded RNA. As more copies of a cloned gene are introduced into a genome, the likelihood is greater that the scenario described in Figure 17.20 will occur. This event silences the expression of the cloned gene even though it is present in multiple copies. Furthermore, if the cloned gene is homologous to a plant gene that is already present in the plant cell (an endogenous gene), this phenomenon will also silence the endogenous plant gene. Therefore, the curious observation made by researchers was that increasing the number of copies of a cloned gene frequently led to gene silencing.

Fire and Mello Show That Double-Stranded RNA Is More Potent Than Antisense RNA at Silencing mRNA

The discovery that double-stranded RNA inhibits mRNA function came about as a result of studies involving gene expression. As a research tool to investigate the function of particular genes, researchers had often introduced antisense RNA (RNA that is complementary to mRNA) into cells as a way to inhibit mRNA translation. Because antisense RNA is complementary to mRNA, the antisense RNA binds to the mRNA, thereby preventing translation. In some experiments, researchers introduced sense RNA (RNA with the same sequence as mRNA) into cells, and, oddly, this also inhibited mRNA translation. Another curious observation was that the effects of antisense RNA often persisted for a very long time, much longer than would have been predicted by the relatively short half-lives of most RNA molecules in the cell. These two unusual observations led Andrew Fire, Craig Mello, and colleagues to investigate how the injection of RNA into cells inhibits mRNA.

They used *C. elegans* as their experimental organism because it was relatively easy to inject with RNA and the expression of many genes had already been established. In 1998, Fire and Mello investigated the effects of several injected RNAs known to be complementary to cellular mRNAs. In the investigation described in **Figure 17.21**, we focus on one of their experiments involving a gene called *mex-3*, which had already been shown to be highly expressed in early embryos. They started with the cloned gene for *mex-3*. This gene was genetically engineered using techniques described in Chapter 19 so that one version had a promoter that would result in the synthesis of the normal mRNA, which is the sense RNA. They made another version of *mex-3* in which a promoter directed the synthesis of the opposite strand, which produced antisense RNA. To make RNA in vitro, they added RNA polymerase and nucleotides to these cloned genes to make sense or antisense RNA.

Next, they injected RNA into the gonads of *C. elegans* hermaphrodites, which was then taken up by the eggs and early embryos. They either injected antisense RNA or they mixed sense and antisense RNA and injected double-stranded RNA. They also analyzed the embryos from uninjected worms as controls. To determine the expression of *mex-3*, they incubated the resulting embryos with a probe complementary to the *mex-3* mRNA. The probe was labeled so it could be observed under the microscope. After this incubation step, any probe that was not bound to this mRNA was washed away.

▶ THE GOAL

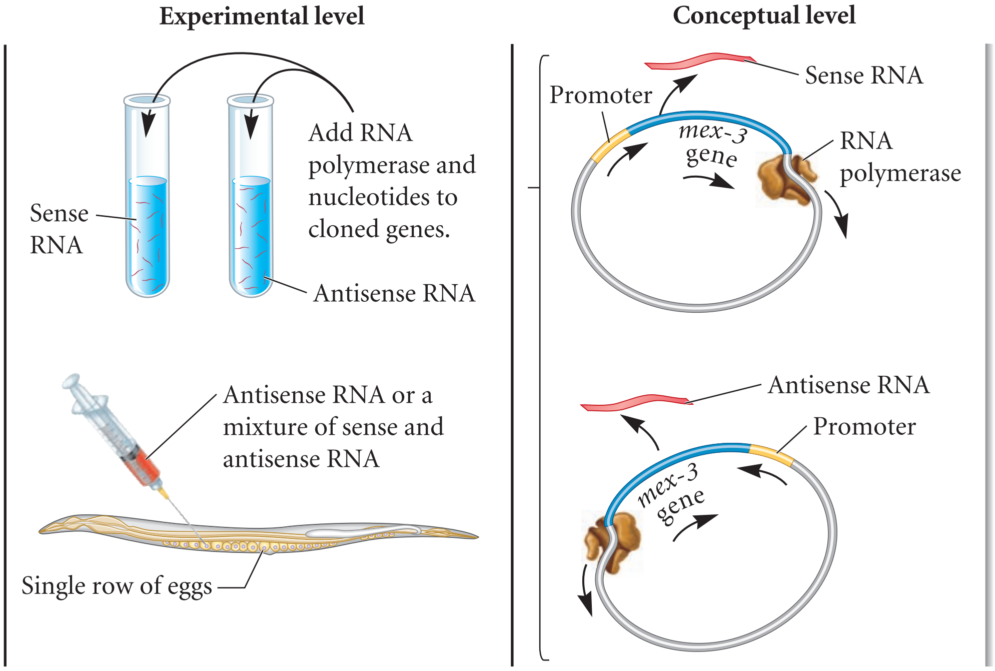
The goal was to further understand how the experimental injection of RNA was responsible for the silencing of particular mRNAs.

▶ **ACHIEVING THE GOAL — FIGURE 17.21 Injection of antisense and double-stranded RNA into *C. elegans* to compare their effects on mRNA silencing.**

Starting material: The researchers used *C. elegans* as their model organism. They also had the cloned *mex-3* gene, which had been previously shown to be highly expressed in the embryo.

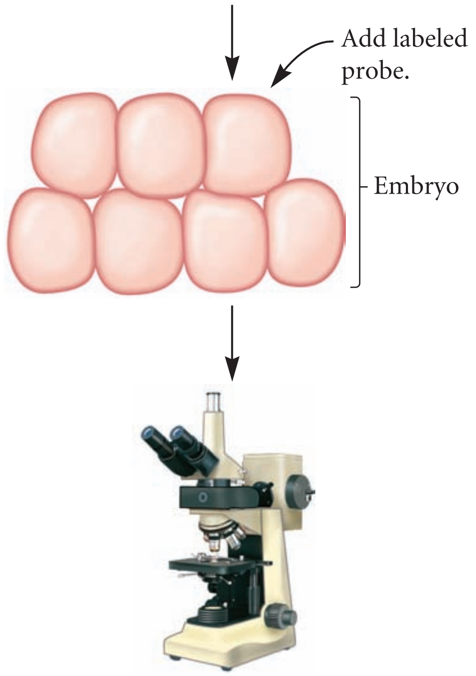
1. Make sense and antisense *mex-3* RNA in vitro using cloned genes for *mex-3* with promoters on either side of the gene. RNA polymerase and nucleotides are added to synthesize the RNAs.

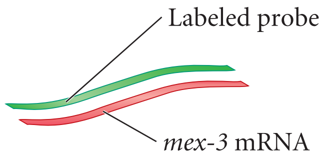
2. Inject either *mex-3* antisense RNA or a mixture of *mex-3* sense and antisense RNA into the gonads of *C. elegans*. This RNA is taken up by the eggs and early embryos. As a control, do not inject any RNA.



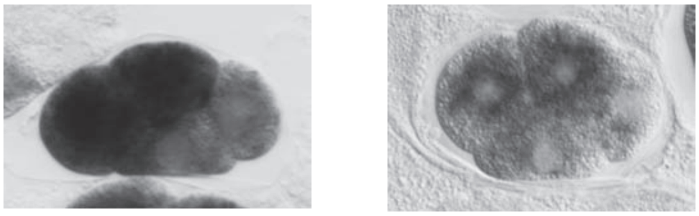
3. Incubate and then subject early embryos to in situ hybridization. In this method, a labeled probe is added that is complementary to *mex-3* mRNA. If cells express *mex-3*, the mRNA in the cells will bind to the probe and become labeled. After incubation with a labeled probe, the cells are washed to remove unbound probe.

4. Observe embryos under the microscope.



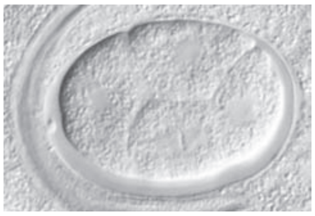


▶ THE DATA



Control

Injected with *mex-3* antisense RNA



Injected with both *mex-3* sense and antisense RNA

Data from A. Fire, S. Xu, M. K. Montgomery, et al. (1998) Potent and specific genetic interference by double-stranded RNA in Caenohabditis elegans. *Nature 391*:6669, 806–811. Photos reprinted by permission of MacMillan Publisher Ltd.

▶ INTERPRETING THE DATA

As seen in the data of Figure 17.21, the control embryos were very darkly stained. These results indicated that the *mex-3* gene is highly expressed, which was known from previous research. In the embryos that had taken up antisense RNA, *mex-3* mRNA levels were decreased, but detectable. Remarkably, in embryos that had taken up double-stranded RNA, no *mex-3* mRNA was detected! These results indicated that double-stranded RNA is more potent at silencing mRNA than is antisense RNA. In this case, the double-stranded RNA caused the *mex-3* mRNA to be degraded. They used the term **RNA interference (RNAi)** to describe the phenomenon in which double-stranded RNA causes the silencing of mRNA. This surprising observation led researchers to investigate the underlying molecular mechanism that accounts for this phenomenon, as described next.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

RNA Interference Is Mediated by MicroRNAs or Short-Interfering RNAs via the RNA-Induced Silencing Complex

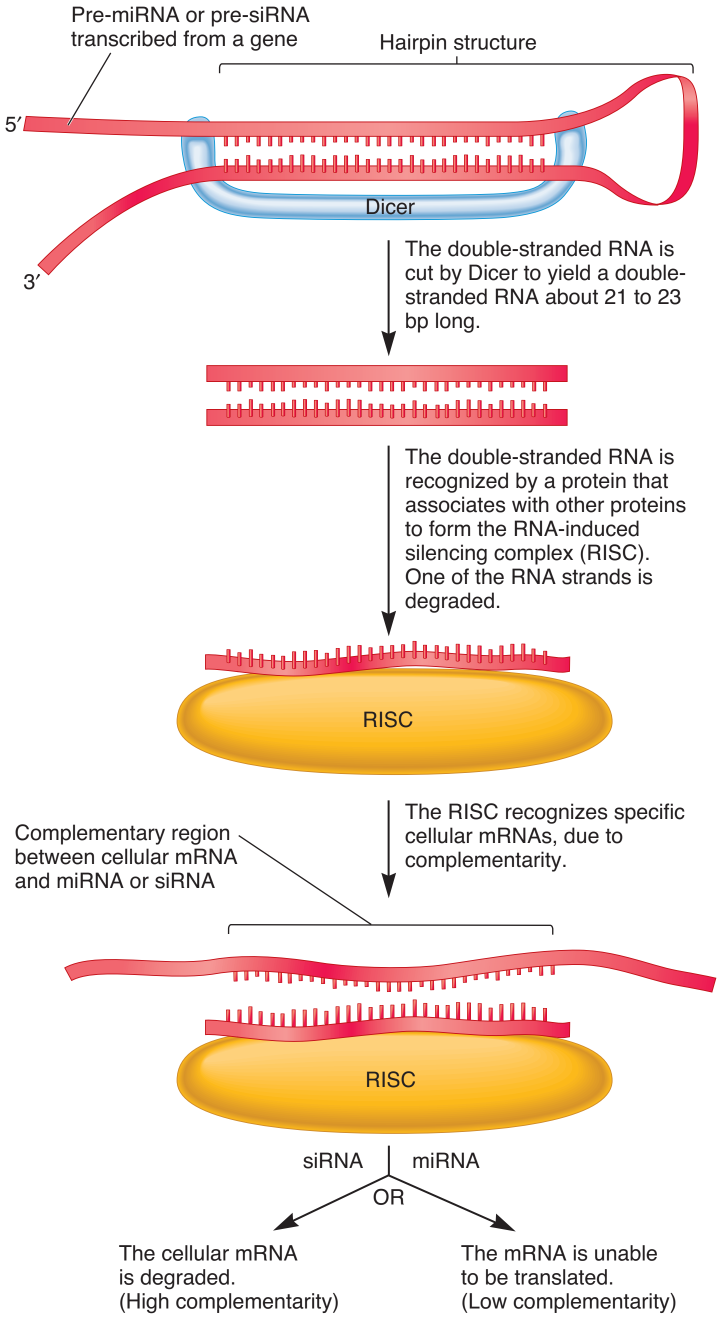
RNA interference is widely found in eukaryotic species. Double-stranded RNA can come from a variety of sources. As mentioned earlier, viruses sometimes produce double-stranded RNA as part of their reproductive cycle, and the insertion of multiple copies of a gene into the genome can also result in double-stranded RNA. In addition, the genomes of eukaryotic organisms have been found to contain genes that encode RNA molecules that cause RNA interference. These RNA molecules, called **microRNAs (miRNAs)** or **short-interfering RNAs (siRNAs),** are small RNA molecules, typically 21 to 23 nucleotides in length, that silence the expression of specific mRNAs. The miRNAs are partially complementary to certain cellular mRNAs and inhibit their translation, whereas siRNAs are usually a perfect match to specific mRNAs and cause the mRNAs to be degraded.

In 1993, Victor Ambros and his colleagues, who were interested in the developmental stages that occur in *C. elegans*, determined that the transcription of a particular gene produced a small RNA, now called a miRNA, that does not encode a protein. Instead, this miRNA was found to be complementary to an mRNA, and it inhibited the translation of the mRNA. Since this study, researchers have discovered that genes encoding miRNAs are widely found in animals and plants. In humans, for example, approximately 200 different genes encode miRNAs.

miRNAs and siRNAs represent an important mechanism of mRNA silencing. How do miRNAs and siRNAs cause the silencing of specific mRNAs? **Figure 17.22** shows how an miRNA or siRNA, encoded by a gene, leads to RNA interference. In this example, the miRNA or siRNA is first synthesized as a pre-miRNA or pre-siRNA, which forms a hairpin structure that is cut by an endonuclease called **Dicer.** This releases double-stranded RNA molecules that associate with cellular proteins to form a complex called the **RNA-induced silencing complex (RISC).** One of the RNA strands is degraded. The remaining single-stranded miRNA or siRNA, which is complementary to specific cellular mRNAs, allows the RISC to specifically recognize and bind to those mRNA molecules.

After binding to the mRNA, two different things may happen. In some cases, the RISC may direct the degradation of the mRNA. This usually occurs when the siRNA and mRNA are a perfect match or are highly complementary. Alternatively, the RISC may inhibit translation. This is more common when the miRNA and mRNA are only partially complementary. In either case, the expression of the mRNA is silenced. The effect is termed RNA interference because the miRNA or siRNA interferes with the proper expression of an mRNA. Likewise, Dicer and RISC may act on double-stranded RNA that comes from viruses or from the insertion of multiple copies of a gene into the genome (as shown at the bottom of Figure 17.20). In 2006, Fire and Mello received the Nobel Prize in physiology or medicine for their discovery of this phenomenon.

RNA interference is believed to have at least three benefits. First, this phenomenon represents a newly identified form of gene regulation. When genes encoding pre-miRNAs or pre-siRNAs are turned on, the production of miRNAs or siRNAs silences the expression of other genes. Second, RNAi may offer a defense mechanism against certain viruses. In particular, this mechanism may inhibit the proliferation of viruses that have a double-stranded RNA genome and viruses that produce double-stranded RNAs as part of their reproductive cycles. After entering the host cell, the viral RNA would be degraded, as in Figure 17.22, and the cell would survive the infection. Third, researchers have speculated that RNAi may also play a role in the silencing of certain transposable elements. As discussed in Chapter 21, transposable elements are DNA segments that have the capacity to move throughout the genome, an event termed *transposition.* Transposable elements carry genes, such as transposase, that are needed in the transposition process. The random insertion of many transposable elements in a cell may ultimately lead to gene silencing due to the scenario described in Figure 17.20. The silencing of genes within transposable elements via RNAi would protect the organism against the potentially harmful effects of transposition.


Pre-miRNA or pre-siRNA transcribed from a gene
Hairpin structure
5′
3′
Dicer
The double-stranded RNA is cut by Dicer to yield a double-stranded RNA about 21 to 23 bp long.
The double-stranded RNA is recognized by a protein that associates with other proteins to form the RNA-induced silencing complex (RISC). One of the RNA strands is degraded.
RISC
The RISC recognizes specific cellular mRNAs, due to complementarity.
Complementary region between cellular mRNA and miRNA or siRNA
RISC
siRNA
miRNA
OR
The cellular mRNA is degraded. (High complementarity)
The mRNA is unable to be translated. (Low complementarity)

ONLINE ANIMATION

FIGURE 17.22 **Mechanism of RNA interference.** A double-stranded region of RNA within a pre-miRNA or pre-siRNA is recognized by an endonuclease called Dicer that digests the RNA to produce a small piece, 21 to 23 bp in length. This small RNA binds with proteins to form an RNA-induced silencing complex (RISC). One strand of the RNA is degraded. The remaining RNA strand in RISC binds to complementary mRNAs. After RISC binding, endonucleases within the complex may cut the mRNA, thereby inactivating it. It should be emphasized that the miRNAs and siRNAs are mRNA-specific. For example, if the miRNA within RISC was complementary to the β-globin mRNA, only the β-globin mRNA would be silenced.

Concept check: Explain why RISC binds to a specific mRNA. What type of bonding occurs?

The Regulation of Iron Assimilation Is an Example of the Regulatory Effect of RNA-Binding Proteins on Translation

In addition to RNA interference, specific mRNAs are sometimes regulated by RNA-binding proteins that directly affect translational initiation or RNA stability. The regulation of iron assimilation provides a well-studied example in which both of these phenomena occur. Before discussing this form of translational control, let's consider the biology of iron metabolism.

Iron is an essential element for the survival of living organisms because it is required for the function of many different enzymes. The pathway by which mammalian cells take up iron is depicted in **Figure 17.23**. Iron (Fe^{3+}) ingested by an animal is absorbed into the bloodstream and becomes bound to transferrin, a protein that carries iron through the bloodstream. The transferrin-Fe^{3+} complex is recognized by a transferrin receptor on the surface of cells. The complex binds to the receptor and then is transported into the cytosol by endocytosis. Once inside, the iron is then released from transferrin. At this stage, Fe^{3+} may bind to cellular enzymes that require iron for their activity. Alternatively, if too much iron is present, the excess iron is stored within a hollow, spherical protein known as ferritin. The storage of excess iron within ferritin helps to prevent the toxic buildup of too much iron within the cell.

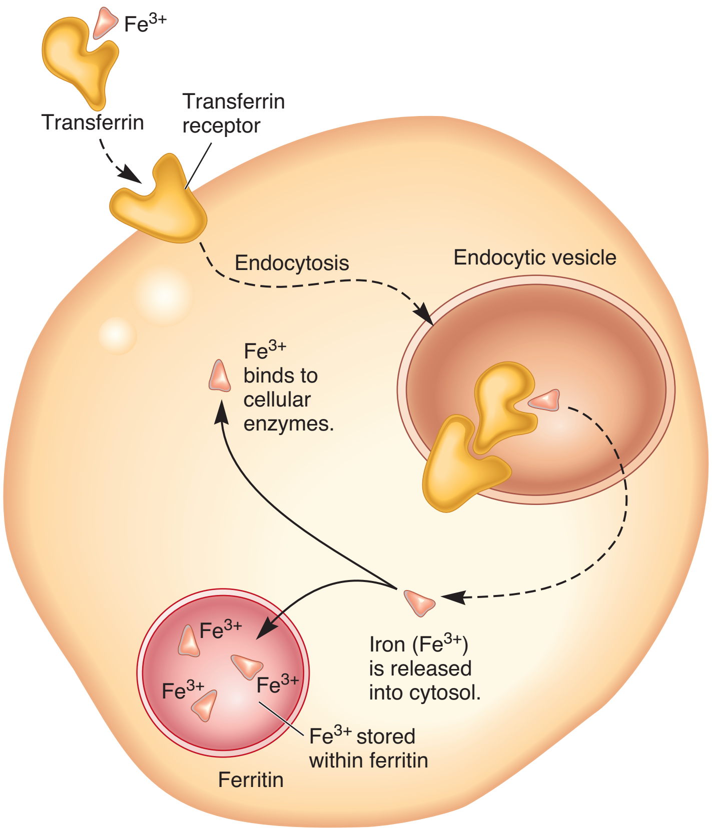

Fe^{3+}
Transferrin
Transferrin receptor
Endocytosis
Endocytic vesicle
Fe^{3+} binds to cellular enzymes.
Iron (Fe^{3+}) is released into cytosol.
Fe^{3+}
Fe^{3+}
Fe^{3+}
Fe^{3+} stored within ferritin
Ferritin

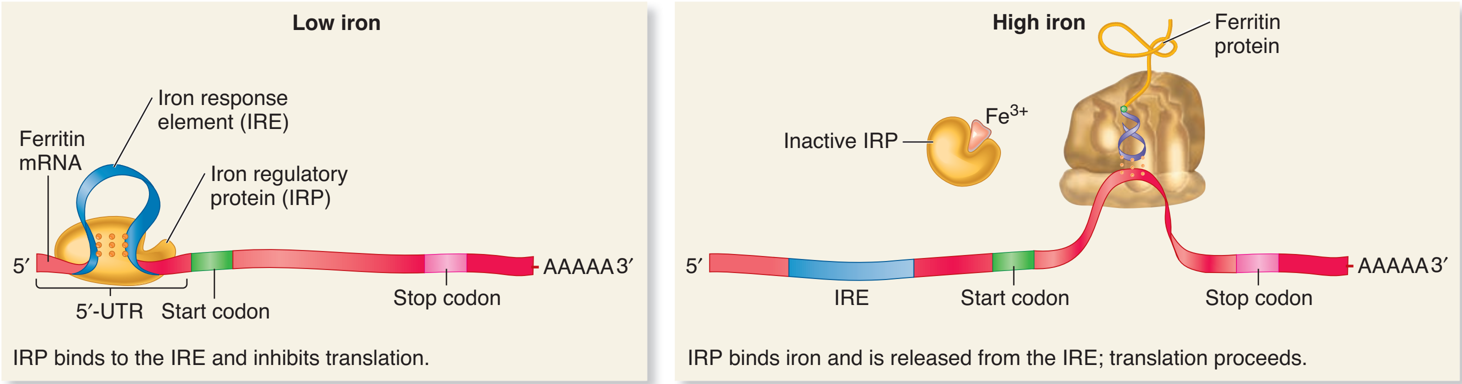
FIGURE 17.23 **The uptake of iron (Fe^{3+}) into mammalian cells.**

Because iron is a vital yet potentially toxic substance, mammalian cells have evolved an interesting way to regulate its assimilation. The two mRNAs that encode ferritin and the transferrin receptor are both influenced by an RNA-binding protein known as the **iron regulatory protein (IRP).** How does IRP exert its effects? This protein binds to a regulatory element within these two mRNAs known as the **iron response element (IRE).** The ferritin mRNA has an IRE in its 5′-untranslated region (5′-UTR). When IRP binds to this IRE, it inhibits the translation of the ferritin mRNA (**Figure 17.24a**, left). However, when iron is abundant in the cytosol, the iron binds directly to IRP and prevents it from binding to the IRE. Under these conditions, the ferritin mRNA is translated to make more ferritin protein (Figure 17.24a, right), which prevents the toxic buildup of iron within the cytosol.

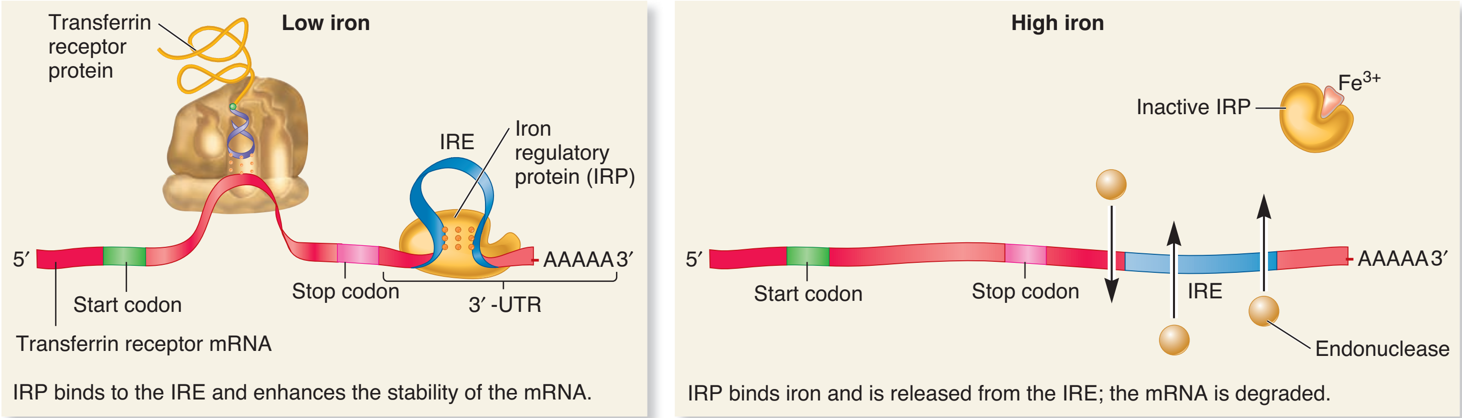
The transferrin receptor mRNA also contains an IRE, but it is located in the 3′-UTR. When IRP binds to this IRE, it does not inhibit translation. Instead, the binding of IRP increases the stability of the mRNA by blocking the action of RNA-degrading enzymes. This leads to increased amounts of transferrin receptor mRNA within the cell when the cytosolic levels of iron are very low (**Figure 17.24b,** left). Under these conditions, more transferrin receptor is made. This promotes the uptake of iron when it is in short supply. In contrast, when iron is abundant within the cytosol, IRP is removed from the transferrin receptor mRNA, and the mRNA becomes rapidly degraded (Figure 17.24b, right). This leads to a decrease in the amount of transferrin receptor, thereby helping to prevent the uptake of too much iron into the cell.

17.5 REVIEWING THE KEY CONCEPTS

- Eukaryotic gene regulation commonly occurs after the mRNA is made. This includes the regulation of RNA processing and translation (see Table 17.2).
- During alternative splicing, proteins called splicing factors regulate which exons are included in a resulting mRNA (see Figures 17.17, 17.18).
- RNA stability may be regulated by an AU-rich element (ARE) found in the 3′-untranslated region (see Figure 17.19).
- The insertion of multiple copies of the same cloned gene into a genome may result in the production of double-stranded RNA that causes mRNA silencing (see Figure 17.20).
- Fire and Mello showed that double-stranded RNA is more potent at inhibiting mRNA than is antisense RNA (see Figure 17.21).
- RNA interference is a mechanism of RNA silencing in which miRNA or siRNA become part of an RNA-induced silencing complex (RISC) that inhibits the translation of mRNA or causes its degradation (see Figure 17.22).
- The regulation of iron uptake and storage prevents the toxic buildup of iron. When iron levels are low, the binding of iron regulatory protein (IRP) to an iron regulatory element (IRE) found in the 5′-untranslated region of ferritin mRNA inhibits translation. By comparison, the binding of IRP to an IRE in the 3′-untranslated region of the transferrin receptor mRNA promotes its stability when iron levels are low (see Figures 17.23, 17.24).


Low iron
Ferritin mRNA
Iron response element (IRE)
Iron regulatory protein (IRP)
5′
AAAAA 3′
5′-UTR
Start codon
Stop codon
IRP binds to the IRE and inhibits translation.
High iron
Ferritin protein
Inactive IRP
Fe3+
5′
AAAAA 3′
IRE
Start codon
Stop codon
IRP binds iron and is released from the IRE; translation proceeds.

(a) Regulation of ferritin mRNA


Transferrin receptor protein
Low iron
IRE
Iron regulatory protein (IRP)
5′
AAAAA 3′
Start codon
Stop codon
3′-UTR
Transferrin receptor mRNA
IRP binds to the IRE and enhances the stability of the mRNA.
High iron
Inactive IRP
Fe3+
5′
AAAAA 3′
Start codon
Stop codon
IRE
Endonuclease
IRP binds iron and is released from the IRE; the mRNA is degraded.

(b) Regulation of transferrin receptor mRNA

FIGURE 17.24 **The regulation of iron assimilation genes by IRP and IRE. (a)** When the Fe^{3+} concentration is low, the binding of the iron regulatory protein (IRP) to the iron response element (IRE) in the 5′-UTR of ferritin mRNA inhibits translation (left). When the Fe^{3+} concentration is high and Fe^{3+} binds to IRP, the IRP is removed from the ferritin mRNA, so translation can proceed (right). **(b)** The binding of IRP to IRE in the 3′-UTR of the transferrin receptor mRNA enhances the stability of the mRNA and leads to a higher concentration of this mRNA. Therefore, more transferrin receptor proteins are made when the Fe^{3+} concentration is low (left). When the Fe^{3+} concentration is high and this metal binds to IRP, the IRP dissociates from the IRE, and the transferrin receptor mRNA is rapidly degraded (right).

Genes→Traits This form of translational control allows cells to use iron appropriately. When the cellular concentration of iron is low, the stability of the transferrin receptor mRNA is increased, thereby enhancing the ability of cells to synthesize more receptors and thereby take up more iron. Also, the translation of ferritin mRNA is inhibited, because ferritin is not needed to store excess iron. By comparison, when the cellular concentration of iron is high, the translation of ferritin mRNA is enhanced. This leads to the synthesis of ferritin and prevents the toxic buildup of iron. When iron is high, the transferrin receptor mRNA is degraded, which decreases further uptake of iron.

Concept check: *If a mutation prevented IRP from binding to the IRE in the ferritin mRNA, how would the mutation affect the regulation of ferritin synthesis? Do you think there would be too much or too little ferritin?*

17.5 COMPREHENSION QUESTIONS

1. What is an advantage of alternative splicing?
   a. Multiple genes can encode the same protein.
   b. The same gene can encode multiple proteins.
   c. The same protein can be digested in multiple ways.
   d. All of the above
2. Which of the following would decrease mRNA stability?
   a. A longer polyA tail, which enhances the binding of polyA-binding protein
   b. A shorter polyA tail, which enhances the binding of polyA-binding protein
   c. A longer polyA tail, which inhibits the binding of polyA-binding protein
   d. A shorter polyA tail, which inhibits the binding of polyA-binding protein
3. The process of RNA interference may lead to
   a. the degradation of an mRNA.
   b. the inhibition of translation of an mRNA.
   c. the synthesis of an mRNA.
   d. both a and b.
4. The binding of iron regulatory protein (IRP) to the iron response element (IRE) in the 5′ region of the ferritin mRNA results in the
   a. inhibition of translation of the ferritin mRNA.
   b. stimulation of translation of the ferritin mRNA.
   c. degradation of the ferritin mRNA.
   d. both a and c.

KEY TERMS

Page 379. gene regulation
Page 380. transcription factor, general transcription factor (GTF), regulatory transcription factors, regulatory elements, control elements, activator, enhancer, repressor, silencer, combinatorial control
Page 381. domain, motif
Page 382. homodimer, heterodimer
Page 383. up regulation, down regulation, orientation-independent, bidirectional, TFIID, coactivators
Page 384. mediator
Page 385. glucocorticoid receptors
Page 386. ATP-dependent chromatin remodeling, chromatin remodeling
Page 387. closed conformation, open conformation, DNA translocases
Page 388. histone variants, histone acetyltransferases
Page 389. histone code hypothesis, nucleosome-free region (NFR)
Page 391. DNA methylation, DNA methyltransferase, CpG islands
Page 392. housekeeping genes, tissue-specific genes, methyl-CpG-binding proteins, de novo methylation, maintenance methylation
Page 394. insulator
Page 396. alternative splicing
Page 397. constitutive exons, alternative exons, splicing factors, SR proteins
Page 398. exon skipping
Page 399. polyA-binding protein, 3′-untranslated region (3′-UTR), AU-rich element (ARE)
Page 401. RNA interference (RNAi), microRNAs (miRNAs), short-interfering RNAs (siRNAs)
Page 402. Dicer, RNA-induced silencing complex (RISC)
Page 403. iron regulatory protein (IRP), iron response element (IRE)
Page 407. promoter bashing

CHAPTER SUMMARY

- Gene regulation refers to the phenomenon that the level of gene expression can be controlled so that genes can be expressed at high or low levels. Gene regulation can occur at any step in the pathway to a functional protein (see Figure 17.1).

17.1 Regulatory Transcription Factors

- Regulatory transcription factors can be activators that bind to enhancers and increase transcription, or they can be repressors that bind to silencers and inhibit transcription (see Figure 17.2).
- Combinatorial control means that the transcription of a gene is controlled by a variety of factors.
- Transcription factor proteins contain specific domains that may be involved in a variety of processes such as DNA binding and protein dimerization (see Figure 17.3).
- Control or response elements are usually orientation-independent.
- Regulatory transcription factors may exert their effects by interacting with TFIID (a general transcription factor) or mediator (see Figures 17.4, 17.5).
- The functions of regulatory transcription factors can be modulated by small effector molecules, protein-protein interactions, and covalent modifications (see Figure 17.6).
- Steroid hormones, such as glucocorticoids, bind to receptors that function as transcriptional activators (see Figure 17.7).

17.2 Chromatin Remodeling, Histone Variation, and Histone Modification

- Chromatin remodeling occurs via ATP-dependent chromatin-remodeling complexes that alter the position and composition of nucleosomes (see Figure 17.8).
- Histone variants, which have amino acid sequences that differ slightly from the standard histones, play specialized roles in chromatin structure and function (see Table 17.1).
- Amino-terminal histone tails are subject to covalent modifications that act as a histone code for the binding of proteins that affect chromatin structure and gene expression (see Figure 17.9).
- Many eukaryotic genes are flanked by a nucleosome-free region (NFR) with well-positioned nucleosomes (see Figure 17.10).
- Gene transcription involves changes in nucleosome locations, composition, and histone modifications (see Figure 17.11).

17.3 DNA Methylation

- DNA methylation is the attachment of a methyl group to a cytosine base via DNA methyltransferase (see Figure 17.12).
- The methylation of CpG islands near promoters usually silences transcription. Methylation may affect the binding of regulatory transcription factors or it may inhibit transcription via methyl-CpG-binding proteins (see Figure 17.13).
- Maintenance methylation can preserve a methylation pattern in daughter cells following cell division (see Figure 17.14).

17.4 Insulators

- An insulator is a segment of DNA that functions as a boundary that protects one gene from the regulatory effects of a neighboring gene. Insulators may act as a barrier to changes in chromatin structure or block the effects of neighboring enhancers (see Figure 17.15).
- Insulators may block the effects of enhancers by promoting the formation of a loop in the DNA (see Figure 17.16).

17.5 Regulation of RNA Processing, RNA Stability, and Translation

- Eukaryotic gene regulation commonly occurs after the mRNA is made. This includes the regulation of RNA processing and translation (see Table 17.2).

- During alternative splicing, proteins called splicing factors regulate which exons are included in a resulting mRNA (see Figures 17.17, 17.18).
- RNA stability may be regulated by an AU-rich element (ARE) found in the 3′-untranslated region (see Figure 17.19).
- The insertion of multiple copies of the same cloned gene into a genome may result in the production of double-stranded RNA that causes mRNA silencing (see Figure 17.20).
- Fire and Mello showed that double-stranded RNA is more potent at inhibiting mRNA than is antisense RNA (see Figure 17.21).
- RNA interference is a mechanism of RNA silencing in which miRNA or siRNA become part of an RNA-induced silencing complex (RISC) that inhibits the translation of mRNA or causes its degradation (see Figure 17.22).
- The regulation of iron uptake and storage prevents the toxic buildup of iron. When iron levels are low, the binding of iron regulatory protein (IRP) to an iron regulatory element (IRE) found in the 5′-untranslated region of ferritin mRNA inhibits translation. By comparison, the binding of IRP to an IRE in the 3′-untranslated region of the transferrin receptor mRNA promotes its stability when iron levels are low (see Figures 17.23, 17.24).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe how the arrangements and compositions of nucleosomes may result in a closed conformation that prevents gene transcription.

Answer: Nucleosome composition and arrangements are important for transcriptional initiation and elongation. The locations of nucleosomes may physically prevent general transcription factors or RNA polymerase from binding to the DNA. Also, nucleosome composition and certain types of covalent modifications may prevent RNA polymerase from forming an open complex, which is necessary to begin transcription. These types of changes are also important for transcription to proceed to the elongation phase.

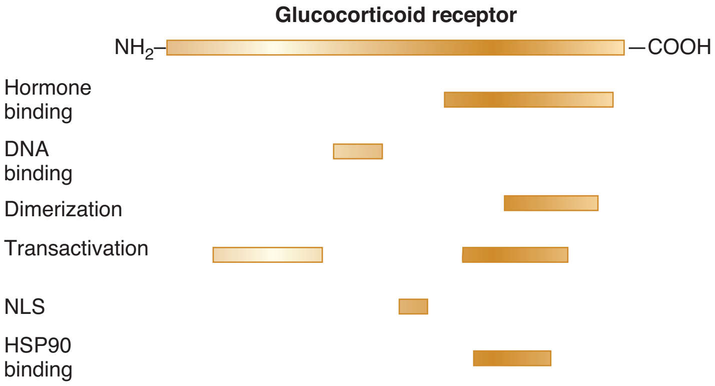
S2. What are the two alternative ways that IRP can affect gene expression at the RNA level?

Answer: The ferritin mRNA has an IRE in its 5′-UTR. When IRP binds to this IRE, it inhibits the translation of the ferritin mRNA. This decreases the amount of ferritin protein, which is not needed when iron levels are low. However, when iron is abundant in the cytosol, the iron binds directly to IRP and prevents it from binding to the IRE. This allows the ferritin mRNA to be translated, producing more ferritin protein. An IRE is also located in the transferrin receptor mRNA in the 3′-UTR. When IRP binds to this IRE, it increases the stability of the mRNA, which leads to an increase in the amount of transferrin receptor mRNA within the cell when the cytosolic levels of iron are very low. Under these conditions, more transferrin receptor is made to promote the uptake of iron, which is in short supply. When iron is found in abundance within the cytosol, IRP is removed from the mRNA, and the mRNA becomes rapidly degraded. This leads to a decrease in the amount of the transferrin receptor.

S3. Eukaryotic regulatory elements are often orientation-independent and can function in a variety of locations. Explain the meaning of this statement.

Answer: Orientation independence means that the regulatory element can function in the forward or reverse orientation. In addition, regulatory elements can function at a variety of locations that may be upstream or downstream from the core promoter. A loop in the DNA must form to bring the regulatory transcription factors bound at the regulatory elements and the core promoter in close proximity to one another.

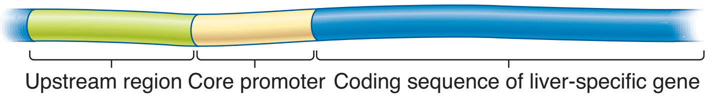
S4. To gain a molecular understanding of how the glucocorticoid receptor works, geneticists have attempted to "dissect" the protein to identify smaller domains that play specific functional roles. Using recombinant DNA techniques described in Chapter 19, particular segments in the coding region of the glucocorticoid receptor gene can be removed. The altered gene can then be expressed in a living cell to see if functional aspects of the receptor have been changed or lost. For example, the removal of the portion of the gene encoding the carboxyl-terminal half of the protein causes a loss of glucocorticoid binding. These results indicate that the carboxyl-terminal portion contains a domain that functions as a glucocorticoid-binding site. The figure shown here illustrates the locations of several functional domains within the glucocorticoid receptor relative to the entire amino acid sequence.



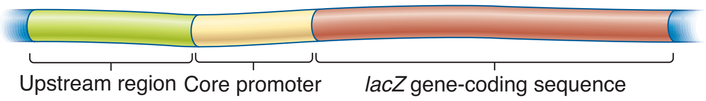
Based on your understanding of the mechanism of glucocorticoid receptor function described in Figure 17.7, explain the functional roles of the different domains in the glucocorticoid receptor.

Answer: The hormone-binding domain is located in the carboxyl-terminal half of the protein. This part of the protein also contains the region necessary for HSP90 binding and receptor dimerization. A nuclear localization sequence (NLS) is located near the center of the protein. After hormone binding, the NLS is exposed on the surface of the protein and allows it to be targeted to the nucleus. The DNA-binding domain, which contains zinc fingers, is also centrally located in the primary amino acid sequence. Zinc fingers promote DNA binding to the major groove. Finally, two separate regions of the protein, one in the amino-terminal half and one in the carboxyl-terminal half, are necessary for the transactivation of RNA polymerase. If these domains are removed from the receptor, it can still bind to the DNA, but it cannot activate transcription.

S5. A common approach to identify genetic sequences that play a role in the transcriptional regulation of a gene is the strategy sometimes called **promoter bashing.** This approach requires gene cloning methods, which are described in Chapter 19. A clone is obtained that has the coding region for a structural gene as well as the region that is upstream from the core promoter. This upstream region is likely to contain genetic regulatory elements such as enhancers and silencers. The diagram shown here depicts a cloned DNA region that contains the upstream region, the core promoter, and the coding sequence for a protein that is expressed in human liver cells. The upstream region may be several thousand base pairs in length.

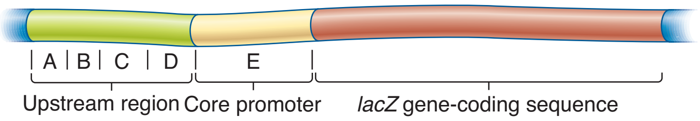


To determine if promoter bashing has an effect on transcription, it is helpful to have an easy way to measure the level of gene expression. One way to accomplish this is to swap the coding sequence of the gene of interest with the coding sequence of another gene. For example, the coding sequence of the *lacZ* gene, which encodes β-galactosidase, is frequently swapped because it is easy to measure the activity of β-galactosidase using an assay for its enzymatic activity. The *lacZ* gene is called a "reporter gene" because its activity is easy to measure. As shown here, the coding sequence of the *lacZ* gene has been swapped with the coding sequence of the liver-specific gene. In this new genetic construct, the expression and transcriptional regulation of the *lacZ* gene is under the control of the core promoter and upstream region of the liver-specific gene.



Now comes the "bashing" part of the experiment. Different segments of the upstream region are deleted, and then the DNA is transformed into living cells. In this case, the researcher would probably transform the DNA into liver cells, because those are the cells where the gene is normally expressed. The last step is to measure the β-galactosidase activity in the transformed liver cells.

In the diagram shown here, the upstream region and the core promoter have been divided into five regions, labeled A–E.



One of these regions was deleted (i.e., bashed out), and the rest of the DNA segment was transformed into liver cells. The data shown here are from this experiment.

Region Deleted	**Percentage of β-Galactosidase Activity***
None	100
A	100
B	330
C	100
D	5
E	<1

*The amount of β-galactosidase activity in the cells carrying an undeleted upstream and promoter region was assigned a value of 100%. The amounts of activity in the cells carrying a deletion were expressed relative to this 100% value.

Explain what these results mean.

Answer: The amount of β-galactosidase activity found in liver cells that do not carry a deletion reflects the amount of expression under normal circumstances. If the core promoter (region E) is deleted, very little expression is observed. This is expected because a core promoter is needed for transcription. If enhancers are deleted, the activity should be less than 100%. It appears that one or more enhancers are found in region D. If a silencer is deleted, the activity should be above 100%. From the data shown, it appears that one or more silencers are found in region B. Finally, if a deletion has no effect, there may not be any regulatory elements there. This was observed for regions A and C.

Note: The deletion of an enhancer has an effect on β-galactosidase activity only if the cell is expressing the regulatory transcription factor that binds to the enhancer and activates transcription. Likewise, the deletion of a silencer has an effect only if the cell is expressing the repressor protein that binds to the silencer and inhibits transcription. In the problem described, the liver cells must be expressing the activator and repressor that recognize the regulatory elements found in regions D and B, respectively.

Conceptual Questions

C1. Discuss the common points of control in eukaryotic gene regulation.

C2. Discuss the structure and function of regulatory elements. Where are they located relative to the core promoter?

C3. What is meant by the term *transcription factor modulation*? List three general ways this can occur.

C4. What are the functions of transcriptional activator proteins and repressor proteins? Explain how they work at the molecular level.

C5. Are the following statements true or false?

A. An enhancer is a type of regulatory element.

B. A core promoter is a type of regulatory element.

C. Regulatory transcription factors bind to regulatory elements.

D. An enhancer may cause the down regulation of transcription.

C6. Transcription factors usually contain one or more motifs that play key roles in their function. What is the function of the following motifs?

A. Helix-turn-helix

B. Zinc finger

C. Leucine zipper

C7. Describe the steps that occur for the glucocorticoid receptor to bind to a GRE.

C8. Let's suppose a mutation in the glucocorticoid receptor does not prevent the binding of the glucocorticoid hormone to the protein but prevents the ability of the receptor to activate transcription. Make a list of all the possible defects that may explain why transcription cannot be activated.

C9. The glucocorticoid receptor is an example of a transcriptional activator that binds to response elements and activates transcription. (Note: The answer to this question is not directly described in this chapter. You have to rely on your understanding of the functioning of other proteins that are modulated by the binding of effector molecules, such as the lac repressor.) How would the function of the glucocorticoid receptor be shut off?

C10. Transcription factors such as the glucocorticoid receptor form homodimers and activate transcription. Other transcription factors form heterodimers. For example, a transcription factor known as myogenic bHLH forms a heterodimer with a protein called the E protein. This heterodimer activates the transcription of genes that promote muscle cell differentiation. However, when myogenic bHLH forms a heterodimer with a protein called the Id protein, transcriptional activation does not occur. (Note: Id stands for Inhibitor of differentiation.) Which of the following possibilities best explains this observation? Only one possibility is correct.

	Myogenic bHLH	**E Protein**	**Id Protein**
Possibility 1			
DNA-binding domain:	Yes	No	No
Leucine zipper:	Yes	No	Yes
Possibility 2			
DNA-binding domain:	Yes	Yes	No
Leucine zipper:	Yes	Yes	Yes
Possibility 3			
DNA-binding domain:	Yes	No	Yes
Leucine zipper:	Yes	No	No

C11. An enhancer, located upstream from a gene, has the following sequence:

5′–GTAG–3′
3′–CATC–5′

This enhancer is orientation-independent. Which of the following sequences also works as an enhancer?

A. 5′–CTAC–3′
3′–GATG–5′

B. 5′–GATG–3′
3′–CTAC–5′

C. 5′–CATC–3′
3′–GTAG–5′

C12. Solved Problem S4 shows the locations of domains in the glucocorticoid receptor relative to the amino- and carboxyl-terminal ends of the protein. Make a drawing that illustrates the binding of a glucocorticoid receptor dimer to the DNA. In your drawing, label the amino- and carboxyl-terminal ends, the hormone-binding domains, the DNA-binding domains, the dimerization domains, and the transactivation domains.

C13. Briefly describe three ways that ATP-dependent chromatin-remodeling complexes may change chromatin structure.

C14. What is a histone variant?

C15. Explain how the acetylation of core histones may loosen chromatin packing.

C16. What is meant by the term *histone code*? With regard to gene regulation, what is the proposed role of the histone code?

C17. What is a nucleosome-free region? Where are they typically found in a genome? How are nucleosome-free regions thought to be functionally important?

C18. Nucleosomes are thought to be displaced as RNA polymerase is transcribing a gene. What would be the potentially harmful consequences if nucleosomes were not put back into a gene after RNA polymerase had passed?

C19. What is an insulator? Describe two different ways that insulators may exert their effects.

C20. What is DNA methylation? When we say that DNA methylation is heritable, what do we mean? How is it passed from a mother to a daughter cell?

C21. Let's suppose that a vertebrate organism carries a mutation that causes some cells that normally differentiate into nerve cells to differentiate into muscle cells. A molecular analysis of this mutation revealed that it was in a gene that encodes a methyltransferase. Explain how an alteration in a methyltransferase could produce this phenotype.

C22. What is a CpG island? Where would you expect one to be located? How does the methylation of CpG islands affect gene expression?

C23. What is the function of a splicing factor? Explain how splicing factors can regulate the tissue-specific splicing of mRNAs.

C24. Figure 17.17 shows the products of alternative splicing for the α-tropomyosin pre-mRNA. Let's suppose that smooth muscle cells produce splicing factors that are not produced in other cell types. Explain where you think such splicing factors bind and how they influence the splicing of the α-tropomyosin pre-mRNA.

C25. Let's suppose a person is homozygous for a mutation in the *IRP* gene that changed the structure of the IRP in such a way that it could not bind iron, but it could still bind to IREs. How would this mutation affect the regulation of ferritin and transferrin receptor mRNAs? Do you think such a person would need more iron in his or her diet than normal individuals? Do you think that excess iron in the diet would be more toxic than it would be for normal individuals? Explain your answer.

C26. In response to potentially toxic substances (e.g., high levels of iron), eukaryotic cells often use translational or posttranslational regulatory mechanisms to prevent cell death, rather than using transcriptional regulatory mechanisms. Explain why.

C27. What are the advantages and disadvantages of mRNAs with a short half-life compared with mRNAs with a long half-life?

C28. What is the phenomenon of RNA interference (RNAi)? During RNAi, explain how the double-stranded RNA is processed and how it leads to the silencing of a complementary mRNA.

C29. With regard to RNAi, what are three possible sources for double-stranded RNA?

C30. What is the relationship between mRNA stability and mRNA concentration? What factors affect mRNA stability?

C31. Describe how the binding of the IRP affects the mRNAs for ferritin and the transferrin receptor. How does iron influence this process?

Application and Experimental Questions

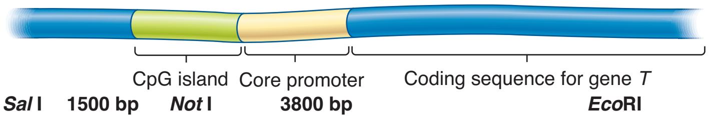
E1. Researchers can isolate a sample of cells, such as skin fibroblasts, and grow them in the laboratory. This procedure is called a cell culture (see Appendix). A cell culture can be exposed to a sample of DNA. If the cells are treated with agents that make their membranes permeable to DNA, the cells may take up the DNA and incorporate the DNA into their chromosomes. This process is called transformation or transfection. Experimenters have transformed human skin fibroblasts with methylated DNA and then allowed the fibroblasts to divide for several cellular generations. The DNA in the daughter cells was then isolated, and the segment that corresponded to the transformed DNA was examined. This DNA segment in the daughter cells was also found to be methylated. However, if the original skin fibroblasts were transformed with unmethylated DNA, the DNA found in the daughter cells was also unmethylated. Do fibroblasts undergo de novo methylation, maintenance methylation, or both? Explain your answer.

E2. Restriction enzymes, described in Chapter 19, are enzymes that recognize a particular DNA sequence and cleave the DNA (along the DNA backbone) at that site. The restriction enzyme known as *Not*I recognizes the sequence

5′–GCGGCCGC–3′
3′–CGCCGGCG–5′

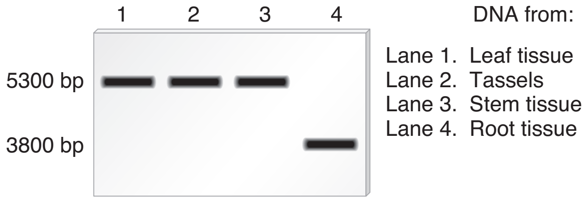
However, if the cytosines in this sequence have been methylated, *Not*I will not cleave the DNA at this site. For this reason, *Not*I is commonly used to investigate the methylation state of CpG islands.

A researcher has studied a gene, which we will call gene *T*, that is found in corn. This gene encodes a transporter involved in the uptake of phosphate from the soil. A CpG island is located near the core promoter of gene *T*. The CpG island has a single *Not*I site. The arrangement of gene *T* is shown here.



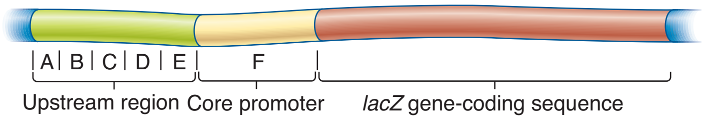
A *Sal*I restriction site is located upstream from the CpG island, and an *Eco*RI restriction site is located near the end of the coding sequence for gene *T*. The distance between the *Sal*I and *Not*I sites is 1500 bp, and the distance between the *Not*I and *Eco*RI sites is 3800 bp. No other sites for *Sal*I, *Not*I, or *Eco*RI are found in this region.

Here is the question. Let's suppose a researcher has isolated DNA samples from four different tissues in a corn plant. These include the leaf, the tassel, a section of stem, and a section of root. The DNA was then digested with all three restriction enzymes, separated by gel electrophoresis, and then probed with a DNA fragment complementary to the gene *T* coding sequence. The results are shown here.



In which type of tissue is the CpG island methylated? Does this make sense based on the function of the protein encoded by gene *T*?

E3. A muscle-specific gene was cloned and then subjected to promoter bashing as described in Solved Problem S5. As shown here, six regions, labeled A–F, were deleted, and then the DNA was transformed into muscle cells.

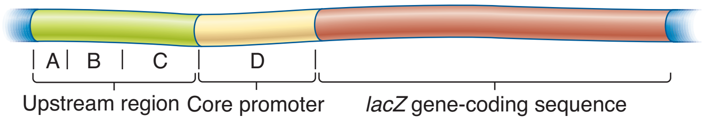


The data shown here are from this experiment.

Region Deleted	Percentage of β-Galactosidase Activity
None	100
A	20
B	330
C	100
D	5
E	15
F	<1

Explain these results.

E4. A gene that is normally expressed in pancreatic cells was cloned and then subjected to promoter bashing as described in Solved Problem S5. As shown here, four regions, labeled A–D, were individually deleted, and then the DNA was transformed into pancreatic cells or into kidney cells.



The data shown here are from this experiment.

Region Deleted	Cell Type Transformed	Percentage of β-Galactosidase Activity
None	Pancreatic	100
A	Pancreatic	5
B	Pancreatic	100
C	Pancreatic	100
D	Pancreatic	<1
None	Kidney	<1
A	Kidney	<1
B	Kidney	100
C	Kidney	<1
D	Kidney	<1

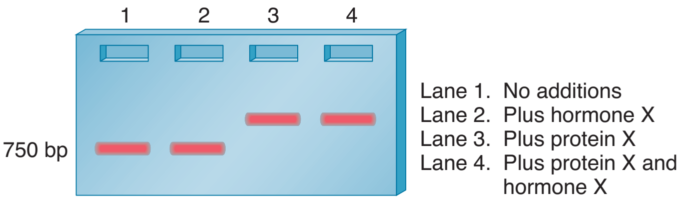
If we assume that the upstream region has one silencer and one enhancer, answer the following questions:

A. Where are the silencer and enhancer located?

B. Why don't we detect the presence of the silencer in the pancreatic cells?

C. Why isn't this gene normally expressed in kidney cells?

E5. A gel retardation assay can be used to determine if a protein binds to a segment of DNA. When a segment of DNA is bound by a protein, its mobility will be retarded, and the DNA band will appear higher in the gel. In the gel retardation assay shown here, a cloned gene fragment that is 750 bp in length contains a regulatory element that is recognized by a transcription factor called protein X. Previous experiments have shown that the presence of hormone X results in transcriptional activation by protein X.

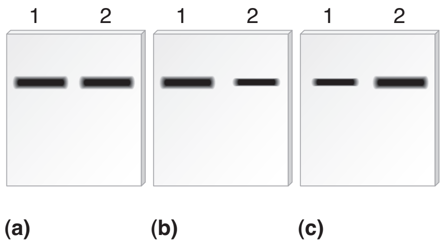


Explain the action of hormone X.

E6. Explain how the data of Fire and Mello suggested that double-stranded RNA is responsible for the silencing of the *mex-3* gene.

E7. Chapter 19 describes a blotting method known as Northern blotting, in which a short segment of cloned DNA is used as a probe to detect RNA that is transcribed from a particular gene. The DNA probe, which is labeled, is complementary to the RNA that the researcher wishes to detect. After the probe DNA binds to the RNA within a blot of a gel, the RNA is visualized as a dark band. The method of Northern blotting can be used to determine the amount of a particular RNA transcribed in a given cell type. If one type of cell produces twice as much of a particular mRNA compared to another cell, the band appears twice as intense.

For this question, a researcher has a DNA probe complementary to the ferritin mRNA. This probe can be used to specifically detect the amount of ferritin mRNA on a gel. A researcher began with two flasks of human skin cells. One flask contained a very low concentration of iron, and the other flask had a high concentration of iron. The mRNA was isolated from these cells and then subjected to Northern blotting, using a probe complementary to the ferritin mRNA. The sample loaded in lane 1 was from the cells grown in a low concentration of iron, and the sample in lane 2 was from the cells grown in a high concentration of iron. Three Northern blots are shown here, but only one of them is correct. Based on your understanding of ferritin mRNA regulation, which blot (a, b, or c) would be your expected result? Explain. Which blot (a, b, or c) would be your expected result if the gel had been probed with a DNA segment complementary to the transferrin receptor mRNA?



Questions for Student Discussion/Collaboration

1. Explain how DNA methylation can regulate gene expression in a tissue-specific way. When and where would de novo methylation occur, and when would demethylation occur? What would occur in the germ line?

2. Enhancers can be almost anywhere and affect the transcription of a gene. Let's suppose you have a gene cloned on a DNA fragment that is 50,000 bp in length. Using cloning methods described in Chapter 19, you can cut out short segments from this 50,000-bp fragment and then reintroduce the smaller fragments into a cell that can express the gene. You would like to know if any enhancers are within the 50,000-bp region that may affect the expression of the gene. Discuss the most efficient strategy you can think of to trim your 50,000-bp fragment, thereby locating enhancers. You can assume that the coding sequence of the gene is in the center of the 50,000-bp fragment and that you can trim the 50,000-bp fragment into any size piece you want using molecular techniques described in Chapter 19.

3. How are regulatory transcription factors and regulatory splicing factors similar in their mechanism of action? In your discussion, consider the domain structures of both types of proteins. How are they different?

Answers to Comprehension Questions

17.1: b, b, b, d
17.2: d, a, b, d
17.3: d, b
17.4: c, c
17.5: b, d, d, a

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

18

The effects of a mutation. *A mutation during embryonic development caused this sheep to have a black spot on its side.*

CHAPTER OUTLINE

18.1 Effects of Mutations on Gene Structure and Function

18.2 Random Nature of Mutations

18.3 Spontaneous Mutations

18.4 Induced Mutations

18.5 DNA Repair

GENE MUTATION AND DNA REPAIR

The primary function of DNA is to store information for the synthesis of cellular proteins. A key aspect of the gene expression process is that the DNA itself does not normally change. This allows DNA to function as a permanent storage unit. However, on relatively rare occasions, a mutation can occur. The term **mutation** refers to a heritable change in the genetic material. This means that the structure of DNA has been changed permanently, and this alteration can be passed from mother to daughter cells during cell division. If a mutation occurs in reproductive cells, it may also be passed from parent to offspring.

The topic of mutation is centrally important in all fields of genetics, including molecular genetics, Mendelian inheritance, and population genetics. Mutations provide the allelic variation that we have discussed throughout this textbook. For example, phenotypic differences, such as tall versus dwarf pea plants, are due to mutations that alter the expression of particular genes. With regard to their phenotypic effects, mutations can be beneficial, neutral, or detrimental. On the positive side, mutations are essential to the continuity of life. They provide the variation that enables species to change and adapt to their environment. Mutations are the foundation for evolutionary change. On the negative side, however, new mutations are much more likely to be harmful rather than beneficial to the individual. The genes within each species have evolved to work properly. They have functional promoters, coding sequences, terminators, and so on, that allow the genes to be expressed. Random mutations are more likely to disrupt these sequences rather than improve their function. For example, many inherited human diseases result from mutated genes. In addition, diseases such as skin and lung cancer can be caused by environmental agents that are known to cause DNA mutations. For these and many other reasons, understanding the molecular nature of mutations is a deeply compelling area of research. In this chapter, we will consider the nature of mutations and their consequences on gene expression at the molecular level.

Because mutations can be quite harmful, organisms have developed several ways to repair damaged DNA. DNA repair systems reverse DNA damage before it results in a mutation that could potentially have negative consequences. DNA repair systems have been studied extensively in many organisms, particularly *Escherichia coli*, yeast, mammals, and plants. A variety of systems repair different types of DNA lesions. We will conclude the chapter by examining the ways that several of these DNA repair systems operate.

18.1 EFFECTS OF MUTATIONS ON GENE STRUCTURE AND FUNCTION

How do mutations affect phenotype? To answer this question, we must appreciate how changes in DNA structure can ultimately affect DNA function. Much of our understanding of mutation has come from the study of experimental organisms, such as bacteria, yeast, and *Drosophila*. Researchers can expose these organisms to environmental agents that cause mutation and then study the consequences of the induced mutations. In addition, because these organisms have a short generation time, researchers can investigate the effects of mutation when they are passed from cell to cell and from parent to offspring.

As discussed in Chapter 8, changes in chromosome structure and number are important occurrences within natural populations of eukaryotic organisms. These types of changes are considered to be mutations because the genetic material has been altered in a way that can be inherited. Changes in chromosome structure and number usually affect the expression of many genes. In comparison, a gene mutation is a relatively small change in DNA structure that affects a single gene. In this section, we will be primarily concerned with the ways that mutations can affect the molecular and phenotypic expression of single genes. We will also consider how the timing of mutations during an organism's development has important consequences.

Gene Mutations Are Molecular Changes in the DNA Sequence of a Gene

A gene mutation occurs when the sequence of the DNA within a gene is altered in a permanent way. It can involve a base substitution in the sequence within a gene or a removal or addition of one or more base pairs.

A **point mutation** is a change in a single base pair within the DNA. For example, the DNA sequence shown here has been altered by a **base substitution,** in which one base is substituted for another.

↓
5′–AACGC**T**AGATC–3′ → 5′–AACGC**G**AGATC–3′
3′–TTGCG**A**TCTAG–5′ 3′–TTGCG**C**TCTAG–5′

A change of a pyrimidine to another pyrimidine, such as C to T, or a purine to another purine, such as A to G, is called a **transition.** This type of mutation is more common than a **transversion,** in which purines and pyrimidines are interchanged. The example just shown is a transversion (a change from T to G), not a transition.

Besides base substitutions, a short sequence of DNA may be deleted from or added to the chromosomal DNA:

(Deletion of 4 bp)
↓
5′–AACGC**TAGA**TC–3′ → 5′–AACGCTC–3′
3′–TTGCG**ATCT**AG–5′ 3′–TTGCGAG–5′

(Addition of 4 bp)
↓
5′–AACGCTAGATC–3′ → 5′–AAC**AGTC**GCTAGATC–3′
3′–TTGCGATCTAG–5′ 3′–TTG**TCAG**CGATCTAG–5′

As we will see next, small deletions or additions to the sequence of a gene can significantly affect the function of the encoded protein.

Gene Mutations Can Alter the Coding Sequence Within a Gene

How might a mutation within the coding sequence of a structural gene affect the amino acid sequence of the polypeptide that is encoded by the gene? **Table 18.1** describes the possible effects of point mutations. **Silent mutations** are those that do not alter the amino acid sequence of the polypeptide even though the nucleotide sequence has changed. Because the genetic code is degenerate, silent mutations can occur in certain bases within a codon, such as the third base, so the specific amino acid is not changed. In contrast, **missense mutations** are base substitutions in which an amino acid change does occur. An example of a missense mutation is the one that causes the human disease known as sickle cell anemia. This disease involves a mutation in the β-globin gene, which alters the polypeptide sequence so the sixth amino acid is changed from a glutamic acid to valine. This single amino acid substitution alters the structure and function of the hemoglobin protein. One consequence of this alteration is that under conditions of low oxygen, the red blood cells assume a sickle shape (**Figure 18.1**). In this case, a single amino acid substitution has a profound effect on the phenotype of cells and even causes a serious disease.

Nonsense mutations involve a change from a normal codon to a stop codon. This terminates the translation of the polypeptide earlier than expected, producing a truncated polypeptide (see Table 18.1). When a nonsense mutation occurs in a bacterial operon, it may also inhibit the expression of downstream genes. This phenomenon, termed **polarity,** is described in Solved Problem S4 at the end of this chapter. Finally, **frameshift mutations** involve the addition or deletion of a number of nucleotides that is not divisible by three. Because the codons are read in multiples of three, this shifts the reading frame. The translation of the mRNA then results in a completely different amino acid sequence downstream from the mutation.

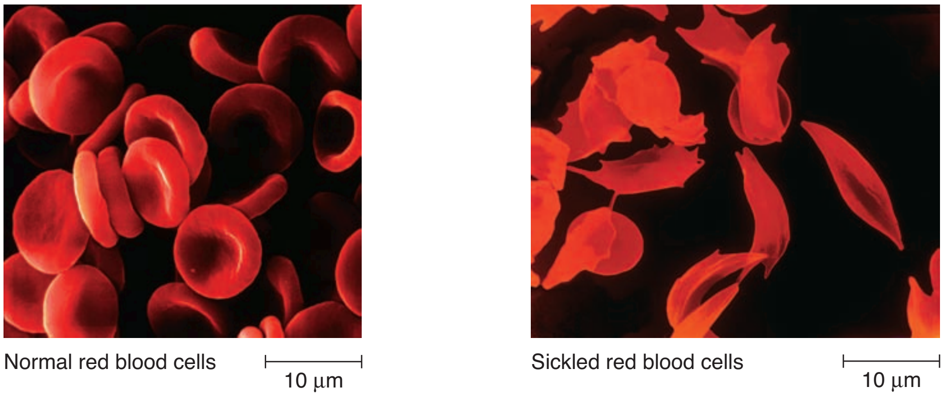
Except for silent mutations, new mutations are more likely to produce polypeptides that have reduced rather than enhanced function. For example, when nonsense mutations produce polypeptides that are substantially shorter, the shorter polypeptides are unlikely to function properly. Likewise, frameshift mutations dramatically alter the amino acid sequence of polypeptides and are thereby likely to disrupt function. Missense mutations are less likely to alter function because they involve a change of a single amino acid within polypeptides that typically contain hundreds of amino acids. When a missense mutation has no detectable effect on protein function, it is referred to as a **neutral mutation.** A missense mutation that substitutes an amino acid with a

TABLE **18.1**

Consequences of Point Mutations Within the Coding Sequence

Type of Change	Mutation in the DNA	Example*	Amino Acids Altered	Likely Effect on Protein Function
None	None	5′–A–T–G–A–C–C–G–A–C–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – Lys – Gly – Thr –	None	None
Silent	Base substitution	↓ 5′–A–T–G–A–C–C–G–A–C–C–C–C–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – Lys – Gly – Thr –	None	None
Missense	Base substitution	↓ 5′–A–T–G–C–C–C–G–A–C–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Pro – Asp – Pro – Lys – Gly – Thr –	One	Neutral or inhibitory
Nonsense	Base substitution	↓ 5′–A–T–G–A–C–C–G–A–C–C–C–G–T–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – STOP!	Many	Inhibitory
Frameshift	Addition/deletion	↓ 5′–A–T–G–A–C–C–G–A–C–G–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Ala – Glu – Arg – Asp –	Many	Inhibitory

*DNA sequence in the coding strand. Note that this sequence is the same as the mRNA sequence except that the RNA contains uracil (U) instead of thymine (T). The 3-base codons are shown in alternating black and red colors. Mutations are shown in green.



(a) Micrographs of red blood cells

NORMAL: NH_2 – VALINE – HISTIDINE – LEUCINE – THREONINE – PROLINE – GLUTAMIC ACID – GLUTAMIC ACID...

SICKLE CELL: NH_2 – VALINE – HISTIDINE – LEUCINE – THREONINE – PROLINE – VALINE – GLUTAMIC ACID...

(b) A comparison of the amino acid sequence between normal β-globin and sickle-cell β-globin

FIGURE 18.1 **Missense mutation in sickle cell anemia. (a)** Normal red blood cells (left) and sickled red blood cells (right). **(b)** A comparison of the amino acid sequence of the normal β-globin polypeptide and the polypeptide encoded by the sickle cell allele. This figure shows only a portion of the polypeptide sequence, which is 146 amino acids long. In β-globin, the first methionine at the amino terminus is removed. As seen here, a missense mutation changes the sixth amino acid from a glutamic acid to a valine.

Genes→Traits A missense mutation alters the structure of β-globin, which is a subunit of hemoglobin, the oxygen-carrying protein in the red blood cells. When an individual is homozygous for this allele, this missense mutation causes the red blood cells to sickle under conditions of low oxygen concentration. The sickling phenomenon is a description of the trait at the cellular level. At the organism level, the sickled cells can clog the capillaries, causing painful episodes, called crises, that can result in organ damage. The shortened life span of the red blood cells leads to symptoms of anemia.

similar chemistry as the original amino acid is likely to be neutral or nearly neutral. For example, a missense mutation that substitutes a glutamic acid for an aspartic acid is likely to be neutral because both amino acids are negatively charged and have similar side chain structures. Silent mutations are also considered a type of neutral mutation.

Mutations can occasionally produce a polypeptide with an enhanced ability to function. Although these favorable mutations are relatively rare, they may result in an organism with a greater likelihood of surviving and reproducing. If this is the case, natural selection may cause such a favorable mutation to increase in frequency within a population. This topic will be discussed later in this chapter and also in Chapter 25.

Gene Mutations Can Occur Outside of the Coding Sequence and Influence Gene Expression

Thus far, we have focused our attention on mutations in the coding regions of genes and their effects on polypeptide structure and protein function. In previous chapters, we learned how various sequences outside the coding sequence play important roles during the process of gene expression. A mutation can occur within noncoding sequences, thereby affecting gene expression (**Table 18.2**). For example, a mutation may alter the sequence within the core promoter of a gene. Promoter mutations that increase transcription are termed **up promoter mutations.** Mutations that make a sequence more like the consensus sequence are likely to be up promoter mutations. In contrast, a **down promoter mutation** causes the promoter to become less like the consensus sequence, decreasing its affinity for transcription factors and decreasing the transcription rate.

TABLE 18.2

Possible Consequences of Gene Mutations Outside of the Coding Sequence

Sequence	Effect of Mutation
Promoter	May increase or decrease the rate of transcription
Regulatory element/operator site	May disrupt the ability of the gene to be properly regulated
5′-UTR/3′-UTR	May alter the ability of mRNA to be translated; may alter mRNA stability
Splice recognition sequence	May alter the ability of pre-mRNA to be properly spliced

In Chapter 16, we considered how mutations could affect regulatory elements. For example, mutations in the *lac* operator site, called *lacO*^{C} mutations, prevent the binding of the lac repressor protein. This causes the *lac* operon to be constitutively expressed even in the absence of lactose. Bacteria strains with *lacO*^{C} mutations are at a selective disadvantage compared with wild-type *E. coli* strains because they waste their energy expressing the *lac* operon even when these proteins are not needed. As noted in Table 18.2, mutations can also occur in other noncoding regions of a gene and alter gene expression in a way that may affect phenotype. For example, mutations that affect the untranslated regions of mRNA—the 5′-UTR and 3′-UTR—may affect gene expression if they alter its ability to be translated or its stability. In addition, mutations in eukaryotic genes can alter splice junctions and affect the order or number of exons contained within mRNA.

Gene Mutations Are Also Given Names That Describe How They Affect the Wild-Type Genotype and Phenotype

Thus far, several genetic terms have been introduced that describe the molecular effects of mutations. Genetic terms are also used to describe the effects of mutations relative to a wild-type genotype or phenotype. In a natural population, a **wild type** is a relatively prevalent genotype. Many or most genes exist as multiple alleles so that a population may have two or more wild-type alleles. A mutation may change a wild-type genotype by altering the DNA sequence of a gene. When such a mutation is rare in a population, it is generally referred to as a **mutant allele.** A reverse mutation, more commonly called a **reversion,** changes a mutant allele back to a wild-type allele.

Another way to describe a mutation is based on its influence on the wild-type phenotype. Mutants are often characterized by their differential ability to survive. As mentioned, a neutral mutation does not alter protein function, so it does not affect survival or reproductive success. A **deleterious mutation,** however, decreases the chances of survival and reproduction. The extreme example of a deleterious mutation is a **lethal mutation,** which results in death to the cell or organism. On the other hand, a **beneficial mutation** enhances the survival or reproductive success of an organism. In some cases, an allele may be either deleterious or beneficial, depending on the genotype and/or the environmental conditions. An example is the sickle cell allele. In the homozygous state, the sickle cell allele lessens the chances of survival. However, an individual who is heterozygous for the sickle cell allele and wild-type allele has an increased chance of survival due to malarial resistance.

Finally, some mutations are called **conditional mutants** because they affect the phenotype only under a defined set of conditions. Geneticists often study conditional mutants in microorganisms; a common example is a temperature-sensitive (*ts*) mutant. A bacterium that has a *ts* mutation grows normally in one temperature range—the permissive temperature range—but exhibits defective growth at a different temperature range—the nonpermissive range. For example, an *E. coli* strain carrying a *ts* mutation may be able to grow from 33°C to 38°C but not between 40°C and 42°C, whereas a wild-type strain can grow at either temperature range.

Suppressor Mutations Reverse the Phenotypic Effects of Another Mutation

A second mutation sometimes affects the phenotypic expression of a first mutation. As an example, let's consider a mutation that causes an organism to grow very slowly. A second mutation at another site in the organism's DNA may restore the normal growth rate, converting the mutant back to the wild-type phenotype. Geneticists call these second-site mutations **suppressors,** or **suppressor mutations.** This name is meant to indicate that a suppressor mutation acts to suppress the phenotypic effects of another mutation. A suppressor mutation differs from a reversion, because it occurs at a DNA site that is distinct from the first mutation.

Suppressor mutations are classified according to their relative locations with regard to the mutation they suppress (**Table 18.3**). When the second mutant site is within the same gene as the first, the mutation is termed an **intragenic suppressor.** This type of suppressor often involves a change in protein structure that compensates for an abnormality in protein structure caused by the first mutation. Researchers often isolate suppressor mutations to obtain information about protein structure and function. For example, Robert Brooker and colleagues have isolated many intragenic suppressors in the *lacY* gene of *E. coli*, which encodes the lactose permease described in Chapter 16. This protein must undergo conformational changes to transport lactose across the cell membrane. They began with single mutations that altered amino acids on transmembrane regions, which inhibited this conformational change, thereby preventing growth on media containing lactose. Suppressor mutations were then isolated that restored transport function and allowed growth on lactose. By analyzing the locations of these suppressor mutations, the researchers were able to determine that certain transmembrane regions in the protein are critical for conformational changes required for lactose transport.

Alternatively, a suppressor mutation can be in a different gene from the first mutation—an **intergenic suppressor.** How do intergenic suppressors work? These suppressor mutations usually involve a change in the expression of one gene that compensates for a loss-of-function mutation affecting another gene (see Table 18.3). For example, a first mutation may cause one protein to be partially or completely defective. An intergenic suppressor mutation in a different structural gene might overcome this defect by altering the structure of a second protein so it could take over the functional role the first protein cannot perform. Alternatively, intergenic suppressors may involve proteins that participate in a common cellular pathway. When a first mutation affects the activity of a protein, a suppressor mutation could alter the function of a second protein involved in this pathway, thereby overcoming the defect in the first.

In some cases, intergenic suppressors involve multimeric proteins, with each subunit encoded by a different gene. A mutation in one subunit that inhibits function may be compensated for by a mutation in another subunit. Another type of intergenic suppressor is one that involves mutations in genetic regulatory proteins such as transcription factors. When a first mutation causes a protein to be defective, a suppressor mutation may occur in a gene that encodes a transcription factor. The mutant transcription factor transcriptionally activates another gene that can compensate for the loss-of-function mutation in the first gene.

Less commonly, intergenic suppression involves mutations in nonstructural genes that alter the translation of particular codons. Examples are suppressor tRNA mutants, which have been identified in bacteria (see Solved Problem S1). Suppressor tRNA mutants have a change in the anticodon region that causes the tRNA to behave contrary to the genetic code. Nonsense suppressors are mutant tRNAs that recognize a stop codon and, instead of stopping translation, put an amino acid into the growing polypeptide chain. This type of mutant tRNA can suppress a nonsense mutation in the mRNA from a structural gene. However, such bacterial strains grow poorly because they may also suppress stop codons in the mRNAs from normal genes.

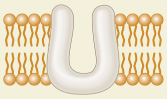
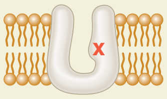
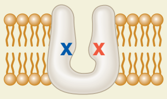
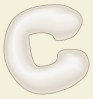
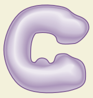
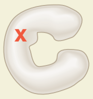
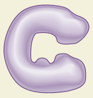
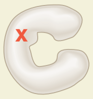
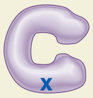
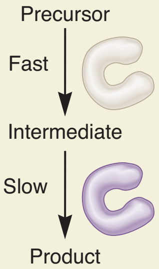
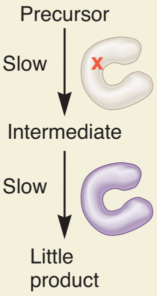
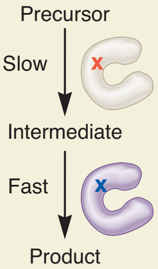
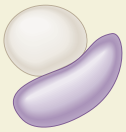
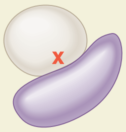
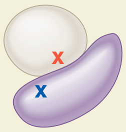
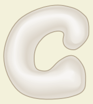
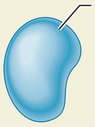
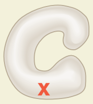
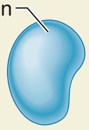
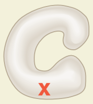
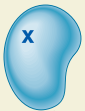
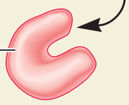
Changes in Chromosome Structure Can Affect the Expression of a Gene

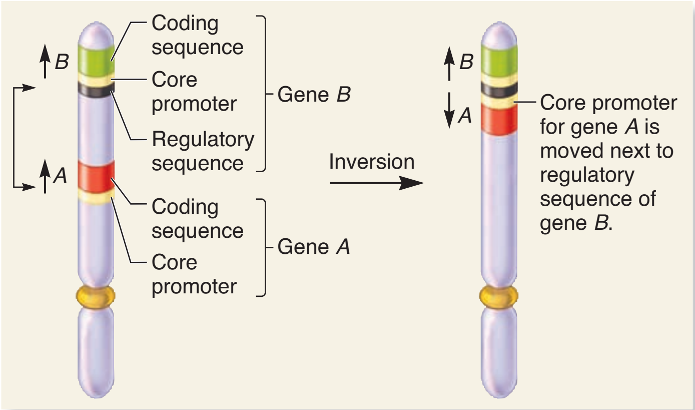
Thus far, we have considered small changes in the DNA sequence of particular genes. A change in chromosome structure can also be associated with an alteration in the expression of single genes. Quite commonly, an inversion or translocation has no obvious phenotypic consequence. However, in 1925, Alfred Sturtevant was the first to recognize that chromosomal rearrangements in *Drosophila melanogaster* can influence phenotypic expression (namely, eye morphology). In some cases, a chromosomal rearrangement may affect a gene because a chromosomal **breakpoint**—the region where two chromosome pieces break and rejoin with other chromosome pieces—occurs within a gene. A breakpoint within the middle of a gene is very likely to inhibit gene function because it separates the gene into two pieces.

In other cases, a gene may be left intact, but its expression may be altered when it is moved to a new location. When this occurs, the change in gene location is said to have a **position effect.** How do position effects alter gene expression? Researchers have discovered two common explanations. **Figure 18.2** depicts a schematic example in which a piece of one chromosome has been inverted or translocated to a different chromosome. One possibility is that a gene may be moved next to regulatory sequences for a different gene, such as silencers or enhancers, that influence the expression of the relocated gene (Figure 18.2a). Alternatively, a chromosomal rearrangement may reposition a gene from a less condensed, or euchromatic chromosome, where it is active, to a very highly condensed, or heterochromatic chromosome. When the gene is moved to a heterochromatic region, its expression may be turned off (Figure 18.2b). This second type of position effect may produce a variegated phenotype in which the expression of the gene is variable. For genes that affect pigmentation, this produces a mottled appearance rather than an even color. **Figure 18.3** shows a position effect that alters eye color in *Drosophila*. Figure 18.3a depicts a normal red-eyed fruit fly, and Figure 18.3b shows a mutant fly that has inherited a chromosomal rearrangement in which a gene affecting eye color has been relocated to a heterochromatic chromosome. The variegated appearance of the eye occurs because the degree of heterochromatin formation varies across different regions of the eye. In cells where heterochromatin formation has turned off the eye color gene, a white phenotype occurs, but other cells allow this same region to remain euchromatic and produce a red phenotype.

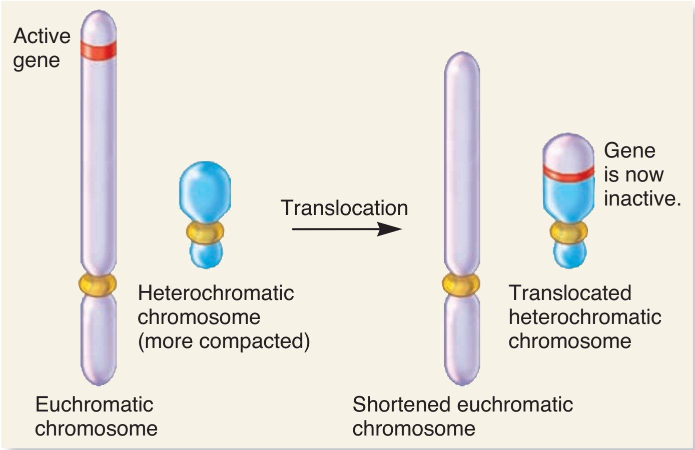
TABLE 18.3

Examples of Suppressor Mutations

Type	No Mutation	First Mutation	Second Mutation	Description
Intragenic	 Transport can occur	 Transport inhibited	 Transport can occur	A first mutation disrupts normal protein function, and a suppressor mutation affecting the same protein restores function. In this example, the first mutation inhibits lactose transport function, and the second mutation restores it.
Intergenic Redundant function	 Enzymatic function 	 Loss of enzymatic function 	  Gain of a new enzymatic function	A first mutation inhibits the function of a protein, and a second mutation alters a different protein to carry out that function. In this example, the proteins function as enzymes.
Common pathway	 Precursor Fast Intermediate Slow Product	 Precursor Slow Intermediate Slow Little product	 Precursor Slow Intermediate Fast Product	Two or more different proteins may be involved in a common pathway. A mutation that causes a defect in one protein may be compensated for by a mutation that alters the function of a different protein in the same pathway.
Multimeric protein	 Active	 Inactive	 Active	A mutation in a gene encoding one protein subunit that inhibits function may be suppressed by a mutation in a gene that encodes a different subunit. The double mutant has restored function.
Transcription factor	 Normal function 	 Loss of function Transcription factor 	 Loss of function Mutant transcription factor turns on a gene that compensates for the loss of function  Causes expression of this protein.  Compensates for inactive protein	A first mutation causes loss of function of a particular protein. A second mutation may alter a transcription factor and cause it to activate the expression of another gene. This other gene encodes a protein that can compensate for the loss of function caused by the first mutation.



(a) Position effect due to regulatory sequences


Active gene
Heterochromatic chromosome (more compacted)
Translocation
Gene is now inactive.
Translocated heterochromatic chromosome
Euchromatic chromosome
Shortened euchromatic chromosome

(b) Position effect due to translocation to a heterochromatic chromosome

FIGURE 18.2 **Causes of position effects.** **(a)** A chromosomal inversion has repositioned the core promoter of gene *A* next to the regulatory sequences for gene *B*. Because regulatory sequences are often bidirectional, the regulatory sequences for gene *B* may regulate the transcription of gene *A*. **(b)** A translocation has moved a gene from a euchromatic to a heterochromatic chromosome. This type of position effect inhibits the expression of the relocated gene.

Concept check: *Explain what the term position effect means.*

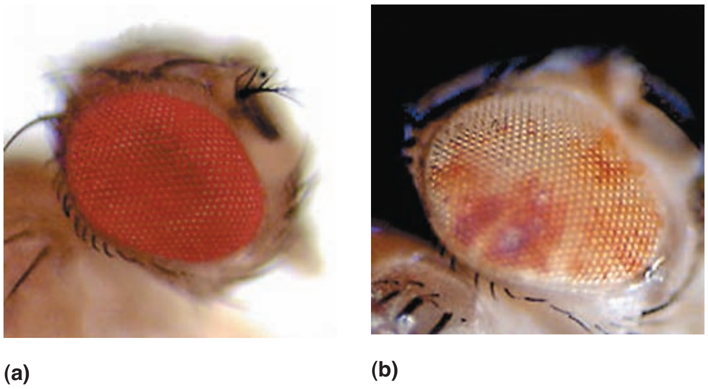


FIGURE 18.3 **A position effect that alters eye color in *Drosophila*.** **(a)** A normal red eye. **(b)** An eye in which an eye color gene has been relocated to a heterochromatic chromosome. This can inactivate the gene in some cells and produces a variegated phenotype.

Genes→Traits Variegated eye color occurs because the degree of heterochromatin formation varies throughout different regions of the eye. In some cells, heterochromatin formation occurs and turns off the eye color gene, thereby leading to the white phenotype. In other patches of cells, the region containing the eye color allele remains euchromatic, yielding a red phenotype.

Concept check: *Has the DNA sequence of the eye color gene been changed in part (b) compared with part (a)? How do we explain the phenotypic difference?*

Mutations Can Occur in Germ-Line or Somatic Cells

In this section, we have considered many different ways that mutations affect gene expression. For multicellular organisms, the timing of mutations also plays an important role. A mutation can occur very early in life, such as in a gamete or a fertilized egg, or it may occur later in life, such as in the embryonic or adult stages. The exact time when mutations occur can be important with regard to the severity of the genetic effect and whether they are passed from parent to offspring.

Geneticists classify the cells of animals into two types: the germ line and the somatic cells. The term **germ line** refers to cells that give rise to the gametes such as eggs and sperm. A **germ-line mutation** can occur directly in a sperm or egg cell, or it can occur in a precursor cell that produces the gametes. If a mutant gamete participates in fertilization, all cells of the resulting offspring will contain the mutation (Figure 18.4a). Likewise, when an individual with a germ-line mutation produces gametes, the mutation may be passed along to future generations of offspring.

The **somatic cells** comprise all cells of the body excluding the germ-line cells. Examples include muscle cells, nerve cells, and skin cells. Mutations can also happen within somatic cells at early or late stages of development. Figure 18.4b illustrates the consequences of a mutation that took place during the embryonic stage. In this example, a **somatic mutation** has occurred within a single embryonic cell. As the embryo grows, this single cell is the precursor for many cells of the adult organism. Therefore, in the adult, a portion of the body contains the mutation. The size of the affected region depends on the timing of the mutation. In general, the earlier the mutation occurs during development, the larger the region. An individual that has somatic regions that are genotypically different from each other is called a **genetic mosaic.**

Figure 18.5 is a photo of an individual with a somatic mutation that occurred during an early stage of development. In this case, the person has a patch of white hair, but the rest of the hair is pigmented. Presumably, this individual initially had a single mutation occur in an embryonic cell that ultimately gave rise to a patch of scalp that produced the white hair. Although a patch of white hair is not a harmful phenotypic effect, mutations during early stages of life can be quite detrimental, especially if they disrupt essential developmental processes. Therefore, even though it is smart to avoid environmental agents that cause

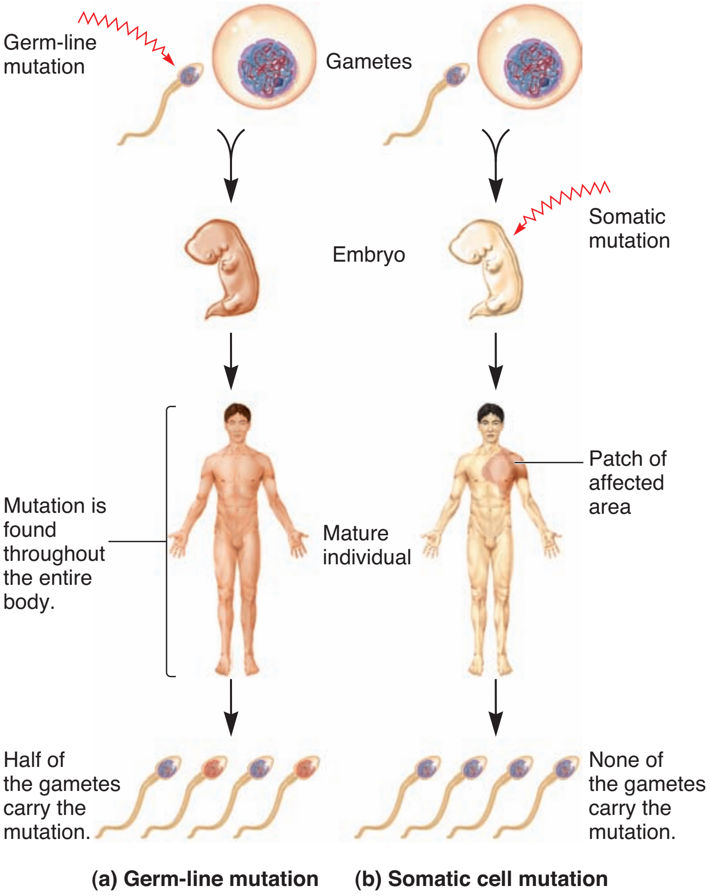
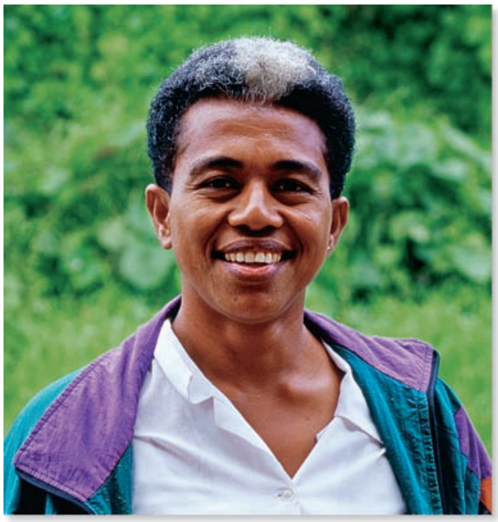

Germ-line mutation
Gametes
Embryo
Somatic mutation
Mutation is found throughout the entire body.
Mature individual
Patch of affected area
Half of the gametes carry the mutation.
None of the gametes carry the mutation.
(a) Germ-line mutation
(b) Somatic cell mutation

FIGURE 18.4 **The effects of germ-line versus somatic mutations.**



© Scott Aitken, www.scottpix.com

FIGURE 18.5 **Example of a somatic mutation.**

Genes→Traits This person has a patch of white hair because a somatic mutation occurred in a single cell during embryonic development that prevented pigmentation of the hair. This cell continued to divide to produce a patch of white hair.

Concept check: *Can this trait be passed to offspring?*

mutations during all stages of life, the possibility of somatic mutations is a rather compelling reason to avoid them during the very early stages of life such as embryonic development, infancy, and early childhood. For example, the possibility of somatic mutations in an embryo is a reason why women are advised to avoid getting X-rays during pregnancy.

18.1 REVIEWING THE KEY CONCEPTS

- A point mutation is a change in a single base pair. These can be transitions or transversions.
- Silent, missense, nonsense, and frameshift mutations may occur within the coding region of a gene (see Table 18.1, Figure 18.1).
- Mutations may also occur within noncoding regions of a gene and affect gene expression (see Table 18.2).
- Suppressor mutations reverse the phenotypic effects of another mutation. They can be intragenic or intergenic (see Table 18.3).
- Changes in chromosome structure can have a position effect that alters gene expression (see Figures 18.2, 18.3).
- With regard to timing, mutations can occur in germ-line or in somatic cells (see Figures 18.4, 18.5).

18.1 COMPREHENSION QUESTIONS

1. A mutation that changes a codon into a stop codon is a
   a. missense mutation.
   b. nonsense mutation.
   c. frameshift mutation.
   d. neutral mutation.
2. A down promoter mutation causes the promoter of a gene to be ________ like the consensus sequence and __________ transcription.
   a. less, stimulates
   b. more, stimulates
   c. less, inhibits
   d. more, inhibits
3. A mutation in one gene that reverses the effects of a mutation in a different gene is
   a. an intergenic suppressor.
   b. an intragenic suppressor.
   c. an up promoter mutation.
   d. a position effect.
4. Which of the following is an example of a somatic mutation?
   a. A mutation in an embryonic muscle cell
   b. A mutation in a sperm cell
   c. A mutation in an adult nerve cell
   d. Both a and c.

18.2 RANDOM NATURE OF MUTATIONS

For a couple of centuries, biologists had questioned whether heritable changes occur purposefully as a result of behavior or exposure to particular environmental conditions or whether they are events that happen randomly. In the nineteenth century, the naturalist Jean-Baptiste Lamarck proposed that physiological events—such as the use or disuse of muscles—determine whether traits are passed along to offspring. For example, his hypothesis suggested that an individual who practiced and became adept at a physical activity and developed muscular legs would pass that characteristic on to the next generation. The alternative point of view is that genetic variation exists in a population as a matter of random chance, and natural selection results in the differential reproductive success of organisms that are better adapted to their environments. Those individuals who, by chance, happen to have beneficial mutations will be more likely to survive and pass these genes to their offspring. These opposing ideas of the nineteenth century—one termed *physiological adaptation* and the other termed *random mutation*—were tested in bacterial studies in the 1940s and 1950s, two of which are described here.

A Fluctuation Test Indicated That Mutations Are Random Events

Salvadore Luria and Max Delbrück were interested in the ability of bacteria to become resistant to infection by a bacteriophage called T1. When a population of *E. coli* cells is exposed to T1, a small percentage of bacteria are found to be resistant to T1 infection and pass this trait to their progeny. Luria and Delbrück were interested in whether such resistance, called *ton*^{r} (T one resistance), is due to the occurrence of random mutations or whether it is a physiological adaptation that occurs at a low rate within the bacterial population.

According to the physiological adaptation hypothesis, the rate of mutation should be a relatively constant value and depend on exposure to the bacteriophage. Therefore, when comparing different populations of bacteria, the number of *ton*^{r} bacteria should be an essentially constant proportion of the total population. In contrast, the random mutation hypothesis depends on the timing of mutation, but not on exposure to the bacteriophage. If a *ton*^{r} mutation occurs early within the proliferation of a bacterial population, many *ton*^{r} bacteria will be found within that population because mutant bacteria will have time to grow and multiply. However, if it occurs much later in population growth, fewer *ton*^{r} bacteria will be observed. In general, a random mutation hypothesis predicts a much greater fluctuation in the number of *ton*^{r} bacteria among different populations.

To distinguish between the physiological adaptation and random mutation hypotheses, Luria and Delbrück inoculated one large flask and 20 individual tubes with *E. coli* cells and grew them in the absence of T1 phage. The flask was grown to produce a very large population of cells, while each individual culture was grown to a smaller population of approximately 20 million cells. They then plated the individual cultures onto media containing T1 phage. Likewise, 10 subsamples, each consisting of 20 million bacteria, were removed from the large flask and plated onto media with T1 phage.

The results of the Luria and Delbrück experiment are shown in **Figure 18.6**. Within the smaller individual cultures, a great fluctuation was observed in the number of *ton*^{r} mutants.

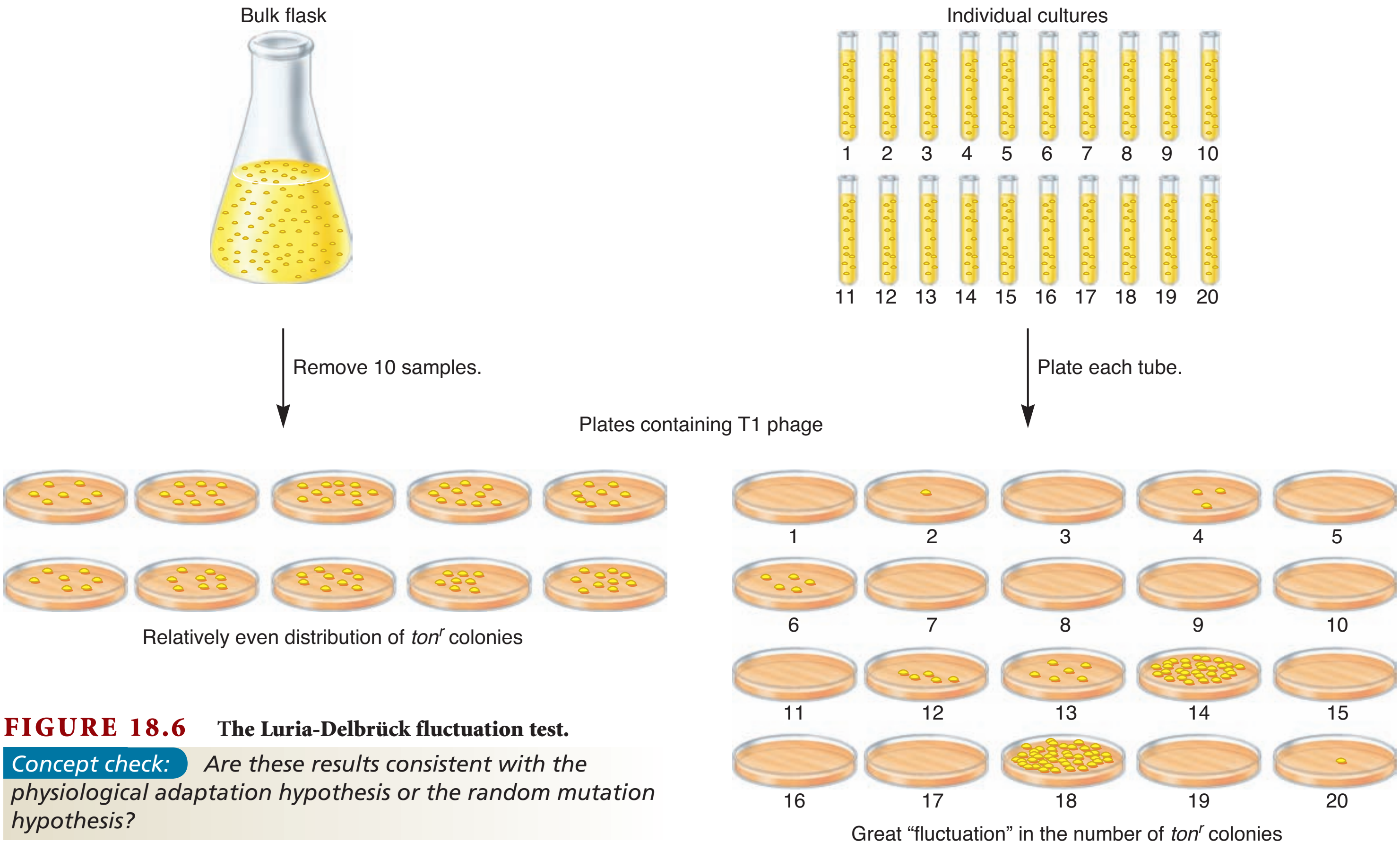


FIGURE 18.6 **The Luria-Delbrück fluctuation test.**

Concept check: Are these results consistent with the physiological adaptation hypothesis or the random mutation hypothesis?

This test, therefore, has become known as the **fluctuation test.** Which hypothesis do these results support? The data are consistent with a random mutation hypothesis in which the timing of a mutation during the growth of a culture greatly affects the number of mutant cells. For example, on plate 14, many *ton*^{r} bacteria were observed. Luria and Delbrück reasoned that a mutation occurred randomly in one bacterium at an early stage of the population growth, before the bacteria were exposed to T1 on plates. This mutant bacterium then divided to produce many daughter cells that inherited the *ton*^{r} trait. In other tubes, such as 1 and 3, this mutation did not occur, so none of the bacteria had a *ton*^{r} phenotype. By comparison, the cells plated from the large flask tended toward a relatively constant and intermediate number of *ton*^{r} bacteria. Because the large flask had so many cells, several independent *ton*^{r} mutations were likely to have occurred during different stages of its growth. In a single flask, however, these independent events would be mixed together to give an average number of *ton*^{r} cells.

Replica Plating Also Revealed that Mutations Are Random Events

Several years later, Joshua and Esther Lederberg were also interested in the relationship between mutation and the environmental conditions under which they occur. To distinguish between the physiological adaptation and random mutation hypotheses, they developed a technique known as **replica plating** in the 1950s. As shown in **Figure 18.7**, they plated a large number of bacteria onto a master plate that did not contain any selective agent (namely, the T1 phage). A sterile piece of velvet cloth was lightly touched to this plate in order to pick up a few bacterial cells from each colony. This replica was then transferred to two secondary plates that contained an agent that selected for the growth of bacterial cells with a particular genotype.

In the example shown in Figure 18.7, the secondary plates contained T1 bacteriophages. On these plates, only those mutant cells that are *ton*^{r} could grow. On the secondary plates, a few colonies were observed. Strikingly, they occupied the same location on each plate. These results indicated that the mutations conferring *ton*^{r} occurred randomly while the cells were growing on the nonselective master plate. The presence of the T1 phage in the secondary plates simply selected for the growth of previously occurring *ton*^{r} mutants. These results supported the random mutation hypothesis. In contrast, the physiological adaptation hypothesis would have predicted that *ton*^{r} bacterial mutants would occur after exposure to the selective agent on the secondary plates. If that had been the case, the colonies would not be expected to arise in identical locations on different secondary plates but rather in random patterns.

Taken together, the results of Luria and Delbrück and those of the Lederbergs supported the random mutation hypothesis, now known as the **random mutation theory.** According to this theory, mutations are a random process—they can occur in any gene and do not involve exposure of an organism to a particular condition that causes specific types of mutations to happen. In some cases, a random mutation may provide a mutant organism with an advantage, such as resistance to T1 phage. Although such mutations occur as a matter of random chance, growth conditions may select for organisms that happen to carry them.

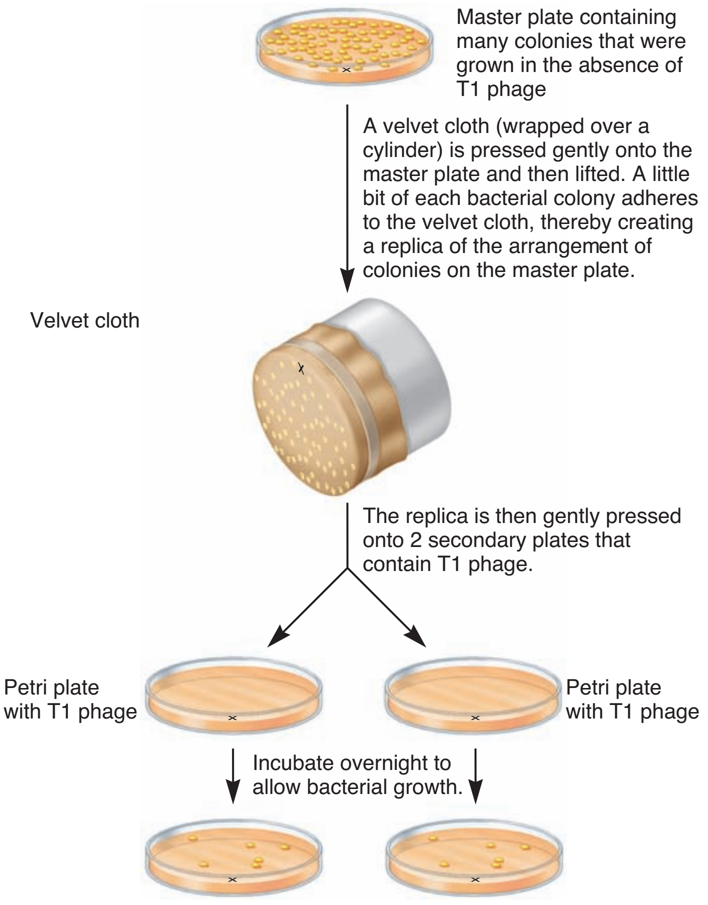

Master plate containing many colonies that were grown in the absence of T1 phage
A velvet cloth (wrapped over a cylinder) is pressed gently onto the master plate and then lifted. A little bit of each bacterial colony adheres to the velvet cloth, thereby creating a replica of the arrangement of colonies on the master plate.
Velvet cloth
The replica is then gently pressed onto 2 secondary plates that contain T1 phage.
Petri plate with T1 phage
Petri plate with T1 phage
Incubate overnight to allow bacterial growth.

FIGURE 18.7 **Replica plating.** Bacteria were first plated on a master plate under nonselective conditions. A sterile velvet cloth was used to make a replica of the master plate. This replica was gently pressed onto two secondary plates that contained a selective agent. In this case, the two secondary plates contained T1 bacteriophage. Only those mutant cells that are *ton*^{r} (resistant to T1) could grow to form visible colonies. Note: The black × indicates the alignment of the velvet and the plates.

Data adapted from J. Lederberg and E. M. Lederberg (1952) Replica plating and indirect selection of bacterial mutants. *J Bacteriol 63,* 399–406.

Concept check: *Why are these results inconsistent with the physiological adaptation hypothesis?*

As researchers have learned more about mutation at the molecular level, the view that mutations are a totally random process has required some modification. Within the same individual, some genes mutate at a much higher rate than other genes. Why does this happen? Some genes are larger than others, which provides a greater chance for mutation. Also, the relative locations of genes within a chromosome may cause some genes to be more susceptible to mutation than others. Even within a single gene, **hot spots** are usually found—certain regions of a gene that are more likely to mutate than other regions (refer back to Chapter 10, Figure 10.13).

18.2 REVIEWING THE KEY CONCEPTS

- The fluctuation test performed by Luria and Delbrück and the replica plating experiments performed by Lederberg and Lederberg were consistent with the random mutation theory (see Figures 18.6, 18.7).

18.2 COMPREHENSION QUESTIONS

1. In the replica plating experiments of the Lederbergs, bacterial colonies appeared at the same locations in each of two secondary plates because
   a. T1 phage caused the mutations to happen.
   b. the mutations occurred on the master plate prior to T1 exposure and prior to replica plating.
   c. both a and b.
   d. none of the above.

18.3 SPONTANEOUS MUTATIONS

Mutations can have a wide variety of effects on the phenotypic expression of genes. For this reason, geneticists have expended a great deal of effort identifying the causes of mutations. This task has been truly challenging, because a staggering number of agents can alter the structure of DNA and thereby cause mutation. Geneticists categorize the cause of mutation in one of two ways. **Spontaneous mutations** are changes in DNA structure that result from natural biological or chemical processes, whereas **induced mutations** are caused by environmental agents (**Table 18.4**). In this section, we will focus on spontaneous mutations and the following section will concern induced mutations.

Many causes of spontaneous mutations are examined in other chapters throughout this textbook. As discussed in Chapter 8, abnormalities in crossing over can produce mutations such as deletions, duplications, translocations, and inversions. In addition, aberrant segregation of chromosomes during meiosis can cause changes in chromosome number. In Chapter 13, we learned that DNA polymerase can make a mistake during DNA replication by putting the wrong base in a newly synthesized daughter strand. Errors in DNA replication are usually infrequent except in certain viruses, such as HIV, that have relatively high rates of spontaneous mutations. Also, normal metabolic processes may produce chemicals within the cell that can react directly with the DNA and alter its structure. As discussed in Chapter 21, transposable elements can alter gene sequences by inserting themselves into genes. In this section, we will examine how spontaneous changes in nucleotide structure, such as depurination, deamination, and tautomeric shifts, can cause mutation. Overall, a distinguishing feature of spontaneous mutations is that their underlying cause originates within the cell. By comparison, the cause of induced mutations originates outside the cell. In this section, we will explore mechanisms, such as depurination, deamination, and tautomeric shifts, which describe how spontaneous mutations can occur, and consider how oxidative stress and trinucleotide repeat expansions cause mutations.

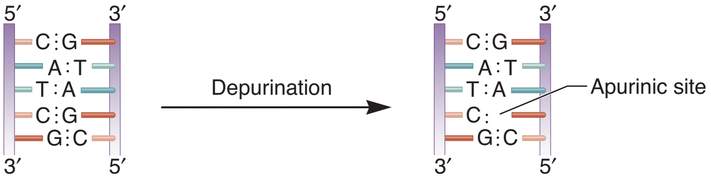
TABLE **18.4**

Causes of Mutations

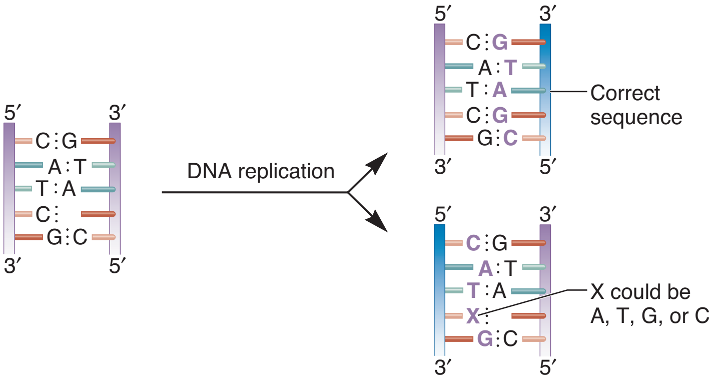
Common Causes of Mutations	Description
Spontaneous	
Aberrant recombination	Abnormal crossing over may cause deletions, duplications, translocations, and inversions (see Chapter 8).
Aberrant segregation	Abnormal chromosomal segregation may cause aneuploidy or polyploidy (see Chapter 8).
Errors in DNA replication	A mistake by DNA polymerase may cause a point mutation (see Chapter 13).
Transposable elements	Transposable elements can insert themselves into the sequence of a gene (see Chapter 21).
Depurination	On rare occasions, the linkage between purines (i.e., adenine and guanine) and deoxyribose can spontaneously break. If not repaired, it can lead to mutation.
Deamination	Cytosine and 5-methylcytosine can spontaneously deaminate to create uracil or thymine.
Tautomeric shifts	Spontaneous changes in base structure can cause mutations if they occur immediately prior to DNA replication.
Toxic metabolic products	The products of normal metabolic processes, such as reactive oxygen species, may be chemically reactive agents that can alter the structure of DNA.
Induced	
Chemical agents	Chemical substances may cause changes in the structure of DNA.
Physical agents	Physical phenomena such as UV light and X-rays can damage the DNA.

Spontaneous Mutations Can Arise by Depurination, Deamination, and Tautomeric Shifts

How can molecular changes in DNA structure cause mutation? Our first examples concern changes that can occur spontaneously, albeit at a low rate. The most common type of chemical change that occurs naturally is **depurination,** which involves the removal of a purine (adenine or guanine) from the DNA. The covalent bond between deoxyribose and a purine base is somewhat unstable and occasionally undergoes a spontaneous reaction with water that releases the base from the sugar, thereby creating an **apurinic site** (**Figure 18.8a**). In a typical mammalian cell, approximately 10,000 purines are lost from the DNA in a 24-hour period at 37°C. The rate of loss is higher if the DNA is exposed to agents that cause certain types of base modifications such as the attachment of alkyl groups (methyl or ethyl groups). Fortunately, as discussed later in this chapter, apurinic sites are recognized by DNA base excision repair enzymes that repair the site. If the repair system fails, however, a mutation may result during subsequent rounds of DNA replication. What happens at an apurinic site during DNA replication? Because a complementary base is not present to specify the



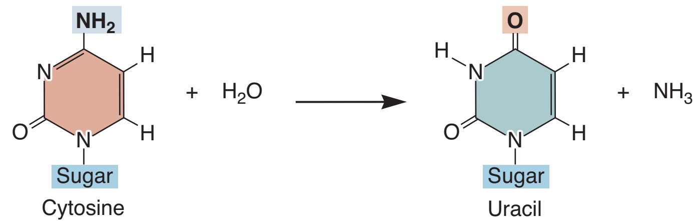
(a) Depurination



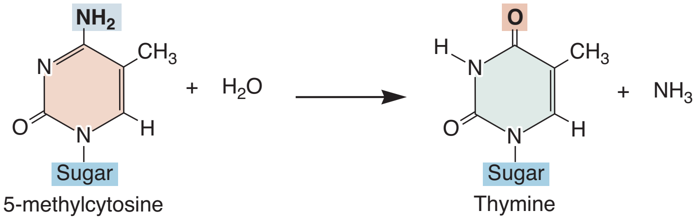
(b) Replication over an apurinic site

FIGURE 18.8 **Spontaneous depurination.** **(a)** The bond between guanine and deoxyribose is broken, thereby releasing the base. This leaves an apurinic site in the DNA. **(b)** If an apurinic site remains in the DNA as it is being replicated, any of the four nucleotides can be added to the newly made strand.

Concept check: When DNA replication occurs over an apurinic site, what is the probability that a mutation will occur?



(a) Deamination of cytosine



(b) Deamination of 5-methylcytosine

FIGURE 18.9 **Spontaneous deamination of cytosine and 5-methylcytosine.** **(a)** The deamination of cytosine produces uracil. **(b)** The deamination of 5-methylcytosine produces thymine.

Concept check: Which of these two changes is more difficult for DNA repair enzymes to fix correctly? Explain why.

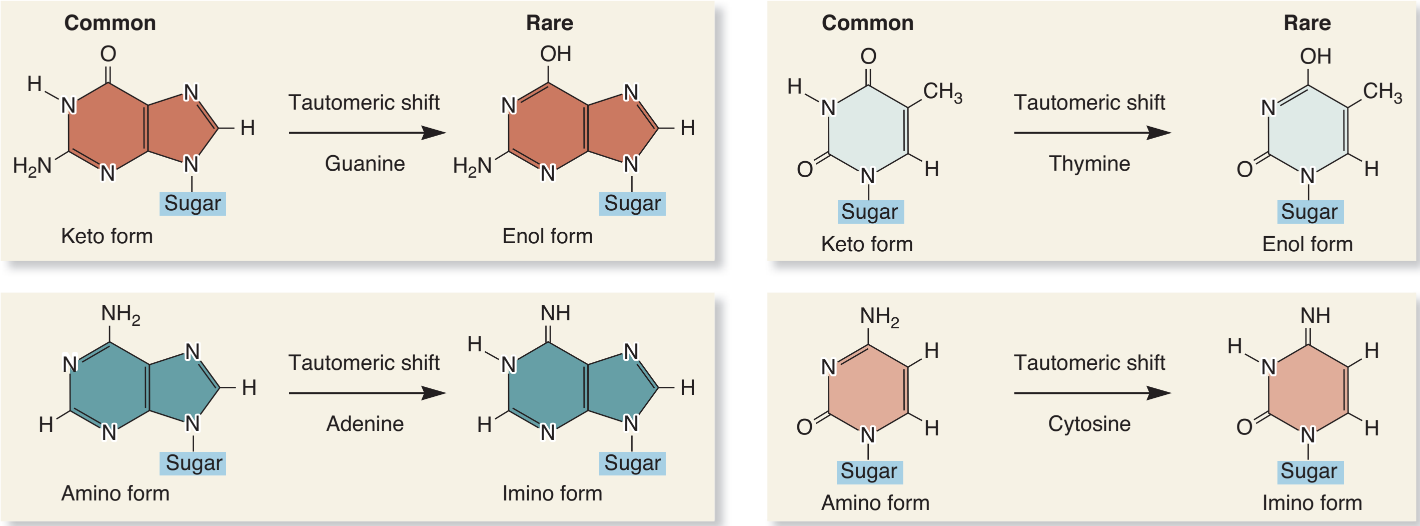
incoming base for the new strand, any of the four bases are added to the new strand in the region that is opposite the apurinic site (**Figure 18.8b**). This may lead to a new mutation.

A second spontaneous lesion that may occur in DNA is the **deamination** of cytosines. The other bases are not readily deaminated. As shown in **Figure 18.9a**, deamination involves the removal of an amino group from the cytosine base. This produces uracil. As discussed later, DNA repair enzymes can recognize uracil as an inappropriate base within DNA and subsequently remove it. However, if such repair does not take place, a mutation may result because uracil hydrogen bonds with adenine during DNA replication. Therefore, if a DNA template strand has uracil instead of cytosine, a newly made strand will incorporate adenine into the daughter strand instead of guanine.

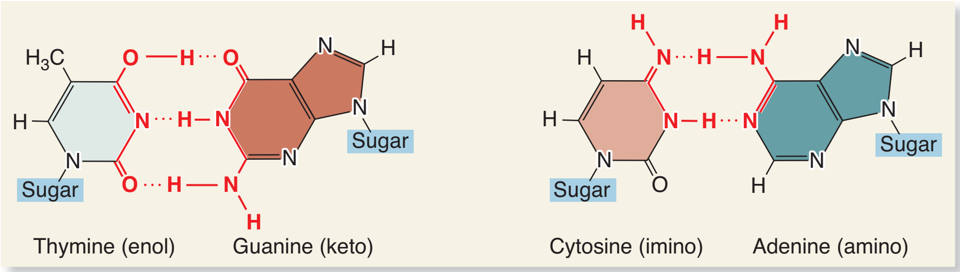
Figure 18.9b shows the deamination of 5-methylcytosine. As discussed in Chapter 17, the methylation of cytosine occurs in many eukaryotic species, forming 5-methylcytosine. It also occurs in prokaryotes. If 5-methylcytosine is deaminated, the resulting base is thymine, which is a normal constituent of DNA. Therefore, this poses a problem for DNA repair because DNA repair proteins cannot distinguish which is the incorrect base—the thymine that was produced by deamination or the guanine in the opposite strand that originally base-paired with the methylated cytosine. For this reason, methylated cytosine bases tend to produce hot spots for mutation. As an example, researchers analyzed 55 spontaneous mutations that occurred within the *lacI* gene of *E. coli* and determined that 44 of them involved changes at sites that were originally occupied by a methylated cytosine base.

A third way that mutations may arise spontaneously involves a temporary change in base structure called a **tautomeric shift.** In this case, the **tautomers** are bases, which exist in keto and enol or amino and imino forms. These forms can interconvert by a chemical reaction that involves the migration of a hydrogen atom and a switch of a single bond and an adjacent double bond. The common, stable form of guanine and thymine is the keto form; the common form of adenine and cytosine is the amino form (**Figure 18.10a**). At a low rate, G and T can interconvert to an enol form, and A and C can change to an imino form. Though the amounts of the enol and imino forms of these bases are relatively small, they can cause a mutation because these rare forms of the bases do not conform to the AT/GC rule of base pairing. Instead, if one of the bases is in the enol or imino form, hydrogen bonding will promote TG and CA base pairs, as shown in **Figure 18.10b**.

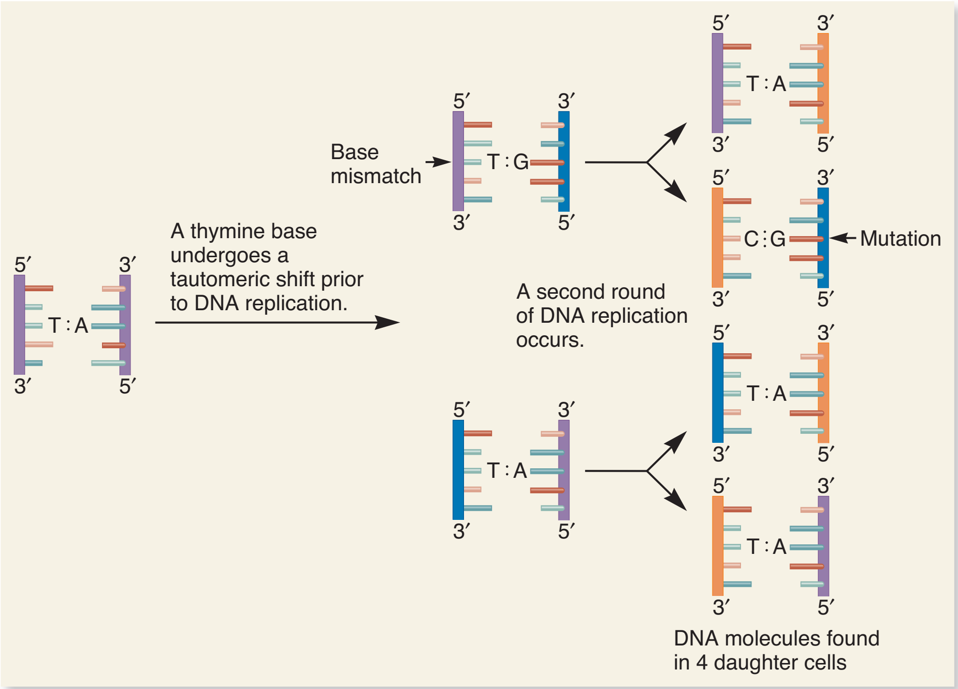
How does a tautomeric shift cause a mutation? The answer is that it must occur immediately prior to DNA replication. When DNA is in a double-stranded condition, the base pairing usually holds the bases in their more stable forms. After the strands unwind, however, a tautomeric shift may occur. In the example shown in **Figure 18.10c**, a thymine base in the template strand has undergone a tautomeric shift just prior to the replication of the complementary daughter strand. During replication, the daughter strand incorporates a guanine opposite this thymine, creating a base mismatch. This mismatch could be repaired via the proofreading function of DNA polymerase or via a mismatch repair


Common
Rare
Tautomeric shift
Guanine
Keto form
Enol form
Thymine
Adenine
Amino form
Imino form
Cytosine
Sugar
H_2N
CH_3
NH_2
OH
NH

(a) Tautomeric shifts that occur in the 4 bases found in DNA


H_3C
Sugar
Thymine (enol)
Guanine (keto)
Cytosine (imino)
Adenine (amino)

(b) Mis–base pairing due to tautomeric shifts


5′
3′
T : A
A thymine base undergoes a tautomeric shift prior to DNA replication.
Base mismatch
T : G
A second round of DNA replication occurs.
C : G
Mutation
DNA molecules found in 4 daughter cells

(c) Tautomeric shifts and DNA replication can cause mutation.

FIGURE 18.10 Tautomeric shifts and their ability to cause mutation. **(a)** The common forms of the bases are shown on the left, and the rare forms produced by a tautomeric shift are shown on the right. **(b)** On the left, the rare enol form of thymine pairs with the common keto form of guanine (instead of adenine); on the right, the rare imino form of cytosine pairs with the common amino form of adenine (instead of guanine). **(c)** A tautomeric shift occurred in a thymine base just prior to replication, causing the formation of a TG base pair. If not repaired, a second round of replication will lead to the formation of a permanent CG mutation. Note: A tautomeric shift is a very temporary situation. During the second round of replication, the thymine base that shifted prior to the first round of DNA replication is likely to have shifted back to its normal form. Therefore, during the second round of replication, an adenine base is found opposite this thymine.

system (discussed later in this chapter). However, if these repair mechanisms fail, the next round of DNA replication will produce a double helix with a CG base pair, whereas the correct base pair should be TA. As shown in the right side of Figure 18.10c, one of four daughter cells inherits this CG mutation.

Oxidative Stress May Also Lead to DNA Damage and Mutation

Aerobic organisms use oxygen as a terminal acceptor of their electron transport chains. **Reactive oxygen species (ROS),** such as hydrogen peroxide, superoxide, and hydroxyl radical, are products of oxygen metabolism in all aerobic organisms. In eukaryotes, ROS are naturally produced as unwanted by-products of energy production in mitochondria. They may also be produced during certain types of immune responses and from a variety of detoxification reactions in the cell. If ROS accumulate, they can damage cellular molecules, including DNA, proteins, and lipids. To prevent this from happening, cells use a variety of enzymes, such as superoxide dismutase and catalase, to prevent the buildup of ROS. In addition, small molecules, such as vitamin C, may act as antioxidants. Likewise, certain foods contain chemicals that act as antioxidants. Colorful fruits and vegetables, including grapes, blueberries, cranberries, citrus fruits, spinach, broccoli, beets, beans, red peppers, carrots, and strawberries, are usually high in antioxidants. In humans, the overaccumulation of ROS has been implicated in a wide variety of medical conditions, including cardiovascular disease, Alzheimer disease, chronic fatigue syndrome, and aging. However, the production of ROS is not always harmful. ROS are produced by the immune system as a means of killing pathogens. In addition, some ROS are used in cell signaling.

Oxidative stress refers to an imbalance between the production of ROS and an organism's ability to break them down. If ROS overaccumulate, one particularly harmful consequence is **oxidative DNA damage,** which refers to changes in DNA structure that are caused by ROS. DNA bases are very susceptible to oxidation. Guanine bases are particularly vulnerable to oxidation, which can lead to several different oxidized guanine products. The most thoroughly studied guanine oxidation product is 7,8-dihydro-8-oxoguanine, which is commonly known as 8-oxoguanine (8-oxoG) (**Figure 18.11**). Researchers often measure the amount of 8-oxoG in a sample of DNA to determine the extent of oxidative stress. Why are oxidized bases harmful? In the case of 8-oxoG, it base pairs with adenine during DNA replication, causing mutations in which a GC base pair becomes a TA base pair. This is a transversion mutation.

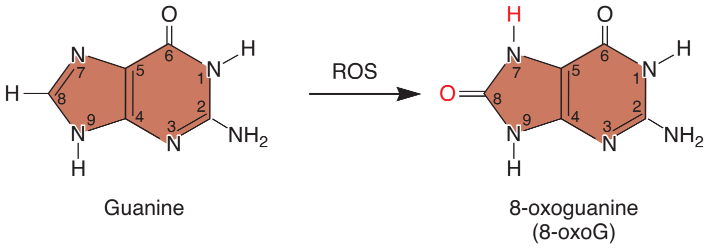


FIGURE 18.11 **Oxidation of guanine to 8-oxoguanine by a reactive oxygen species (ROS).**

Concept check: *What is a reactive oxygen species?*

Although oxidative DNA damage can occur spontaneously, it also results from environmental agents, such as ultraviolet light, X-rays, and many chemicals, including those found in cigarette smoke. Later in this chapter, we will discuss such environment agents and their abilities to cause mutation.

DNA Sequences Known as Trinucleotide Repeats Are Hotspots for Mutation

Researchers have discovered that several human genetic diseases are caused by an unusual form of mutation known as **trinucleotide repeat expansion (TNRE).** The term refers to the phenomenon in which a repeated sequence of three nucleotides can readily increase in number from one generation to the next. In humans and other species, certain genes and chromosomal locations contain regions where trinucleotide sequences are repeated in tandem. These sequences are usually transmitted normally from parent to offspring without mutation. However, in persons with TNRE disorders, the length of a trinucleotide repeat has increased above a certain critical size and thereby causes disease symptoms.

Table 18.5 describes several human diseases that involve these types of expansions: spinal and bulbar muscular atrophy (SBMA), Huntington disease (HD), spinocerebellar ataxia (SCA1), fragile X syndromes (FRAXA and FRAXE), and myotonic muscular dystrophy (also called dystrophia myotonica, DM). In some cases, the expansion is within the coding sequence of the gene. Typically, such an expansion is a CAG repeat. Because CAG encodes glutamine, these repeats cause the encoded proteins to contain long tracts of glutamine. The presence of glutamine tracts causes the proteins to aggregate. This aggregation of proteins or protein fragments carrying glutamine repeats is correlated with the progression of the disease. In other TNRE disorders, the expansions are located in the noncoding regions of genes. In the case of the two fragile X syndromes, the repeat produces CpG islands that become methylated. As discussed in Chapter 17, DNA methylation can silence gene transcription. For DM, it has been hypothesized that these expansions cause abnormal changes in RNA structure, which produce disease symptoms.

Some TNRE disorders have the unusual feature of a progressively worsening severity in future generations—a phenomenon called **anticipation** or **dynamic mutation.** An example is depicted here, where the trinucleotide repeat of CAG has expanded from 11 tandem copies to 18.

CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG $n = 11$

to

CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
GCAGCAGCAGCAG $n = 18$

However, anticipation does not occur with all TNRE disorders and usually depends on whether the disease is inherited

TABLE 18.5
TNRE Disorders

Disease	SBMA	HD	SCA1	FRAXA	FRAXE	DM
Repeat Sequence	CAG	CAG	CAG	CGG	GCC	CTG
Location of Repeat	Coding sequence	Coding sequence	Coding sequence	5′-UTR	5′-UTR	3′-UTR
Number of Repeats in Unaffected Individuals	11–33	6–37	6–44	6–53	6–35	5–37
Number of Repeats in Affected Individuals	36–62	27–121	43–81	>200	>200	>200
Pattern of Inheritance	X-linked	Autosomal dominant	Autosomal dominant	X-linked	X-linked	Autosomal dominant
Disease Symptoms	Neuro-degenerative	Neuro-degenerative	Neuro-degenerative	Mental impairment	Mental impairment	Muscle disease
Anticipation*	None	Male	Male	Female	None	Female

*Indicates the parent in which anticipation occurs most prevalently.

SBMA, spinal and bulbar muscular atrophy; HD, Huntington disease; SCA1, spinocerebellar ataxia; FRAXA and FRAXE, fragile X syndromes; DM, dystrophia myotonica (myotonic muscular dystrophy).

from the mother or father. In the case of HD, anticipation is likely to occur if the mutant gene is inherited from the father. In contrast, DM is more likely to get worse if the gene is inherited from the mother. These results suggest that TNRE happens more frequently during oogenesis or spermatogenesis, depending on the particular gene involved. The phenomenon of anticipation makes it particularly difficult for genetic counselors to advise couples about the likely severity of these diseases if they are passed to their children.

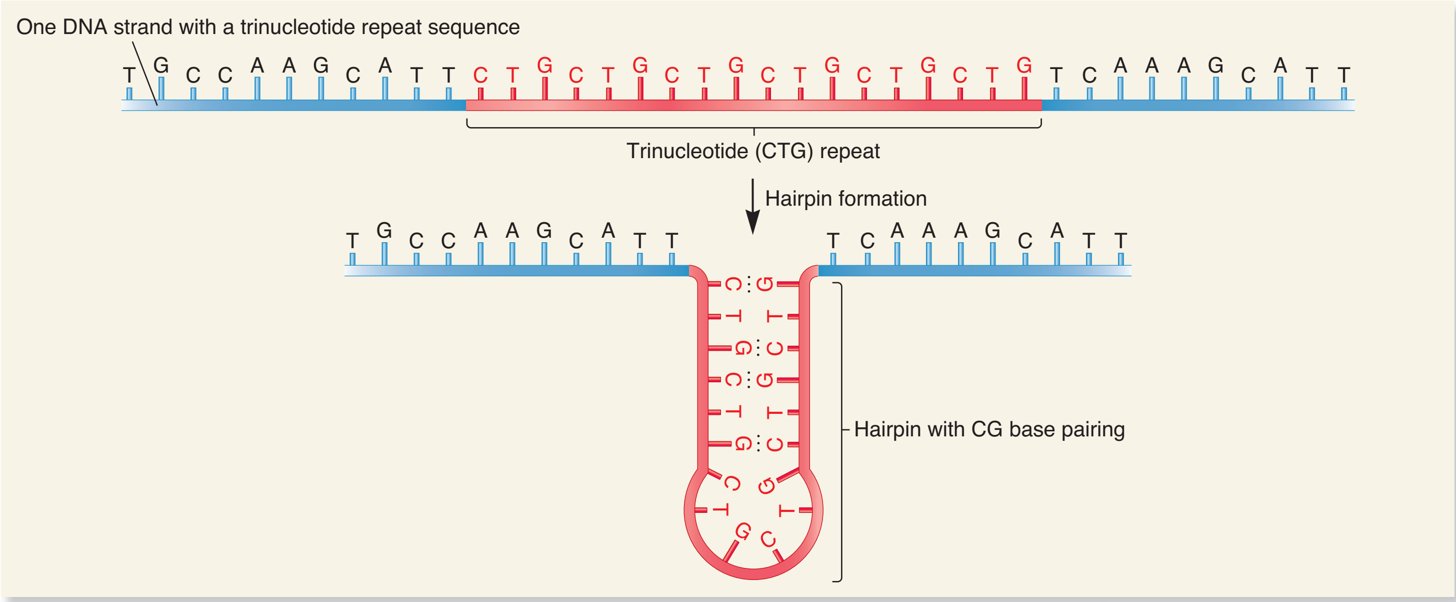
How does TNRE occur? Though it may occur in more than one way, researchers have determined that a key aspect of TNRE is that the triplet repeat can form a hairpin, also called a stem-loop. A consistent feature of the triplet sequences associated with TNRE is they contain at least one C and one G (see Table 18.5). As shown in **Figure 18.12a**, such a sequence can form a hairpin due to the formation of CG base pairs. The formation of a hairpin during DNA replication can lead to an increase in the length of a DNA region if it occurs in the newly made daughter strand (**Figure 18.12b**). In such a scenario, DNA polymerase slips off the DNA after the repeat sequence is synthesized. A hairpin quickly forms, and then DNA polymerase hops back onto the DNA and continues with DNA replication. When this occurs, DNA polymerase is synthesizing most of the hairpin region twice. Depending on how the DNA is repaired, this may result in a TNRE. Alternatively, if a hairpin forms in the template strand, a deletion may result in the trinucleotide repeat, which is also common (**Figure 18.12c**). When the trinucleotide repeat sequence is abnormally long, such changes may frequently occur during gamete formation, and therefore future generations may have trinucleotide repeat sequences that are longer (or shorter) than their parents. In addition, as discussed later in this chapter, certain forms of DNA repair involve short regions where the DNA is replicated. Such DNA repair may occur in the somatic cells of individuals who already have abnormally long trinucleotide repeats in a particular gene, which may frequently cause those repeats to become even longer and more harmful. In this way, the severity of some TNRE disorders may get worse with age.

18.3 REVIEWING THE KEY CONCEPTS

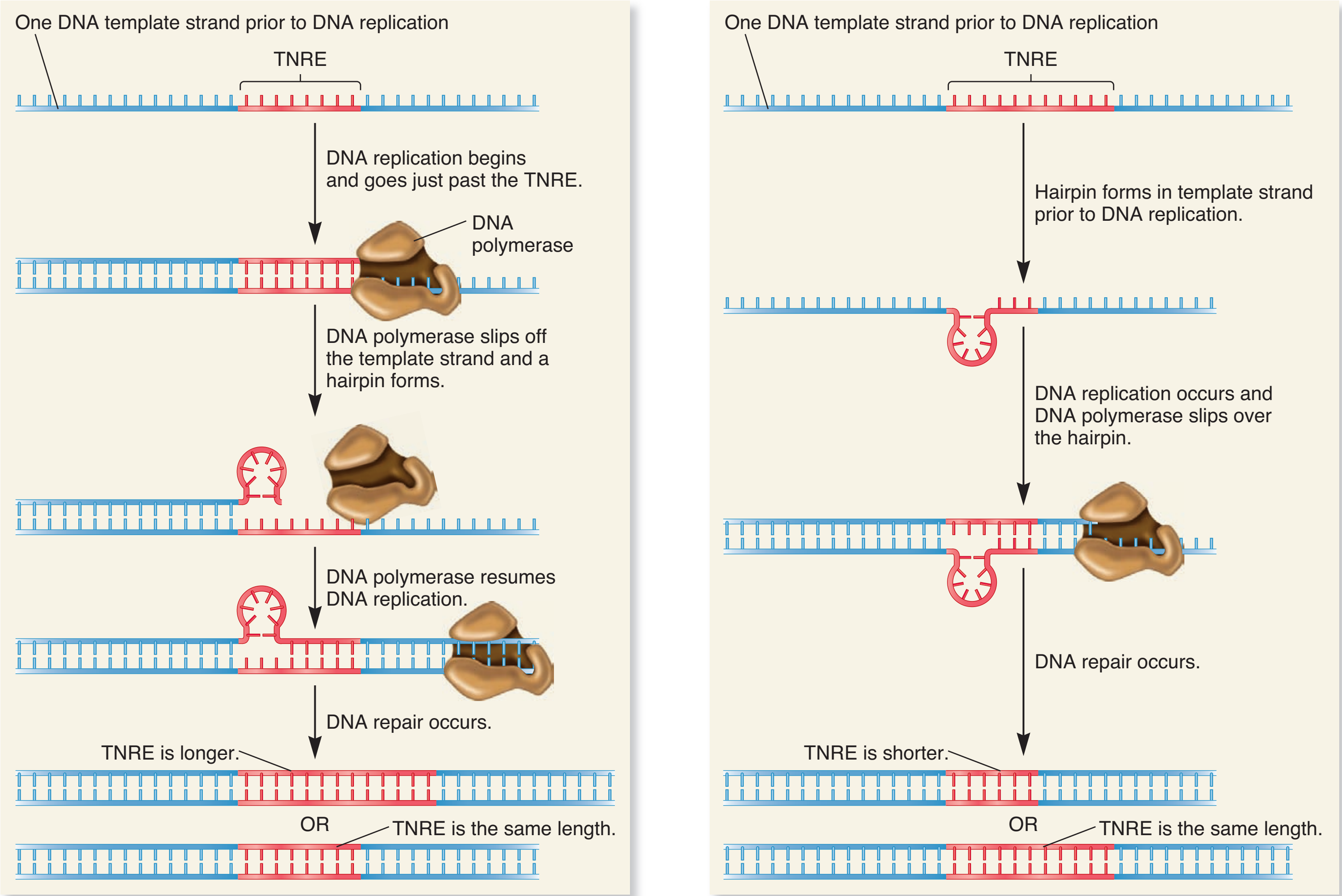
- Spontaneous mutations result from natural biological and chemical processes, whereas induced mutations are caused by environmental agents (see Table 18.4).
- Three common ways that mutations can arise spontaneously is by depurination, deamination, and tautomeric shifts (see Figures 18.8–18.10).
- Reactive oxygen species (ROS) can cause spontaneous mutations by oxidizing bases in DNA (see Figure 18.11).
- In individuals with trinucleotide repeat expansion (TNRE), a trinucleotide repeat increases above a certain critical size and becomes prone to frequent expansion. This type of mutation, called a dynamic mutation, is responsible for certain types of human diseases. The repeats can expand due to hairpin formation during DNA replication (see Table 18.5, Figure 18.12).

18.3 COMPREHENSION QUESTIONS

1. Which of the following is *not* an example of a spontaneous mutation?
   a. A mutation caused by an error in DNA replication
   b. A mutation caused by a tautomeric shift
   c. A mutation caused by UV light
   d. None of the above
2. A point mutation could be caused by
   a. depurination.
   b. deamination.
   c. tautomeric shift.
   d. all of the above.
3. One way that TNRE may occur involves the formation of __________ that disrupts _____________.
   a. a double-strand break, chromosome segregation
   b. an apurinic site, DNA replication
   c. a hairpin, DNA replication
   d. a free radical, DNA structure


One DNA strand with a trinucleotide repeat sequence
T G C C A A G C A T T C T G C T G C T G C T G C T G C T G T C A A A G C A T T
Trinucleotide (CTG) repeat
Hairpin formation
T G C C A A G C A T T
T C A A A G C A T T
Hairpin with CG base pairing

(a) Formation of a hairpin with a trinucleotide (CTG) repeat sequence


One DNA template strand prior to DNA replication
TNRE
DNA replication begins and goes just past the TNRE.
DNA polymerase
DNA polymerase slips off the template strand and a hairpin forms.
DNA polymerase resumes DNA replication.
DNA repair occurs.
TNRE is longer.
OR
TNRE is the same length.
One DNA template strand prior to DNA replication
TNRE
Hairpin forms in template strand prior to DNA replication.
DNA replication occurs and DNA polymerase slips over the hairpin.
DNA repair occurs.
TNRE is shorter.
OR
TNRE is the same length.

(b) Mechanism of trinucleotide repeat expansion

(c) Mechanism of trinucleotide repeat deletion

FIGURE 18.12 **Proposed mechanism of trinucleotide repeat expansion (TNRE).** **(a)** Trinucleotide repeats can form hairpin structures due to CG base pairing. **(b)** Formation of a trinucleotide repeat expansion. **(c)** Formation of a deletion.

18.4 INDUCED MUTATIONS

As we have seen, spontaneous mutations can occur in a wide variety of ways. They result from natural biological and chemical processes. In this section, we turn our attention to induced mutations that are caused by environmental agents. These agents can be either chemical or physical. They enter the cell and lead to changes in DNA structure. Agents known to alter the structure of DNA which lead to mutations are called **mutagens.** In this section, we will explore several mechanisms by which mutagens can alter the structure of DNA. We will also consider laboratory tests that can identify potential mutagens.

Mutagens Alter DNA Structure in Different Ways

Over the past few decades, researchers have found that an enormous array of agents act as mutagens that permanently alter the structure of DNA. We often hear in the news media that we should avoid these agents in our foods and living environment. For example, we use products such as sunscreens to help us avoid the mutagenic effects of ultraviolet (UV) rays. The public is concerned about mutagens for two important reasons. First, mutagenic agents are often involved in the development of human cancers. In addition, because new mutations may be deleterious, people want to avoid mutagens to prevent gene mutations that may have harmful effects on their future offspring.

Mutagenic agents are usually classified as chemical or physical mutagens. Examples of both types of agents are listed in Table 18.6. In some cases, chemicals that are not mutagenic can be altered to a mutagenically active form after being ingested. Cellular enzymes such as oxidases have been shown to activate some mutagens.

Mutagens alter the structure of DNA in various ways. Some mutagens act by covalently modifying the structure of bases. For example, **nitrous acid** (HNO_2) replaces amino groups with keto groups ($=NH_2$ to $=O$), a process called deamination. This changes cytosine to uracil and adenine to hypoxanthine. When this altered DNA replicates, the modified bases do not pair with the appropriate bases in the newly made strand. Instead, uracil pairs with adenine, and hypoxanthine pairs with cytosine (Figure 18.13).

Other chemical mutagens also disrupt the appropriate pairing between nucleotides by alkylating bases within the DNA. During alkylation, methyl or ethyl groups are covalently attached to the bases. Examples of alkylating agents include **nitrogen mustard** (a type of mustard gas) and **ethyl methanesulfonate (EMS).** Mustard gas was used as a chemical weapon during World War I. Such agents severely damage the skin, eyes, mucous membranes, lungs, and blood-forming organs.

Some mutagens exert their effects by directly interfering with the DNA replication process. For example, **acridine dyes** such as **proflavin** contain flat structures that intercalate, or insert themselves, between adjacent base pairs, thereby distorting the helical structure. When DNA containing these mutagens is replicated, single-nucleotide additions and/or deletions can occur in the newly made daughter strands, creating a frameshift mutation.

TABLE 18.6
Examples of Mutagens

Mutagen	Effect(s) on DNA Structure
Chemical	
Nitrous acid	Deaminates bases
Nitrogen mustard	Alkylating agent
Ethyl methanesulfonate	Alkylating agent
Proflavin	Intercalates within DNA helix
5-Bromouracil	Base analog
2-Aminopurine	Base analog
Physical	
X-rays	Cause base deletions, single-stranded nicks in DNA, crosslinking, and chromosomal breaks
UV light	Promotes pyrimidine dimer formation, such as thymine dimers

Compounds such as **5-bromouracil (5BU)** and **2-aminopurine** are base analogs that become incorporated into daughter strands during DNA replication. 5BU is a thymine analog that can be incorporated into DNA instead of thymine. Like thymine, 5BU base-pairs with adenine. However, at a relatively high rate, 5BU undergoes a tautomeric shift and base-pairs with guanine (Figure 18.14a). When this occurs during DNA replication, 5BU causes a mutation in which a TA base pair is changed to a 5BU-G base pair (Figure 18.14b). This is a transition, because the adenine has been changed to a guanine, both of which are purines. During the next round of DNA replication, the template strand containing the guanine base produces a GC base pair. In this way, 5BU promotes a change of an AT base pair into a GC base pair.

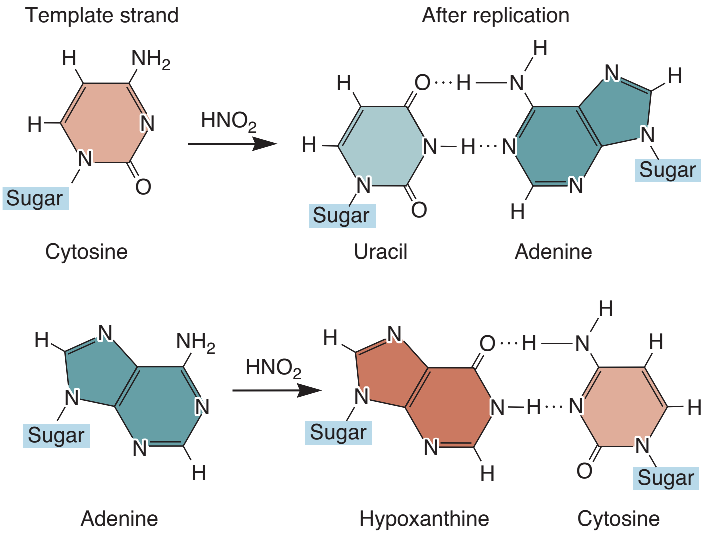
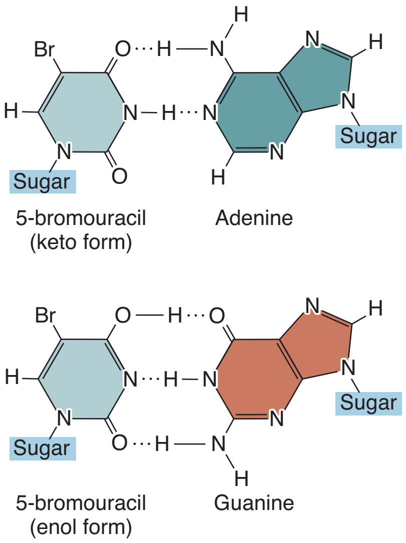
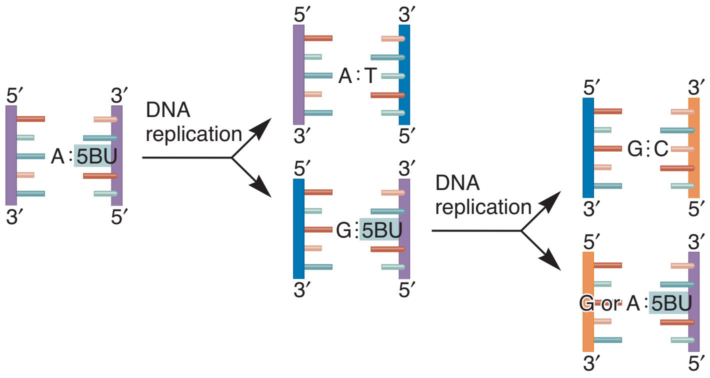


FIGURE 18.13 Mispairing of modified bases that have been deaminated by nitrous acid. Nitrous acid converts cytosine to uracil, and adenine to hypoxanthine. During DNA replication, uracil pairs with adenine, and hypoxanthine pairs with cytosine. This creates mutations in the newly replicated strand during DNA replication.


Br
O···H—N
H
N
H
N
H—
N—H···N
Sugar
N
Sugar
O
N
H
5-bromouracil
(keto form)
Adenine
Br
O—H···O
N
H
H—
N···H—N
N
Sugar
N
N
Sugar
O···H—N
H
5-bromouracil
(enol form)
Guanine

(a) Base pairing of 5BU with adenine or guanine


5′
3′
A : 5BU
3′
5′
DNA
replication
5′
3′
A : T
3′
5′
5′
3′
G : 5BU
3′
5′
DNA
replication
5′
3′
G : C
3′
5′
5′
3′
G or A : 5BU
3′
5′

(b) How 5BU causes a mutation in a base pair during DNA replication

FIGURE 18.14 **Base pairing of 5-bromouracil and its ability to cause mutation.** **(a)** In its keto form, 5BU bonds with adenine; in its enol form, it bonds with guanine. **(b)** During DNA replication, guanine may be incorporated into a newly made strand by pairing with 5BU. After a second round of replication, the DNA contains a GC base pair instead of the original AT base pair.

Concept check: *Does 5-bromouracil cause a transition or a transversion?*

Compounds like 5BU are sometimes used in chemotherapy for cancer. The rationale is that these compounds are incorporated only into the DNA of actively dividing cells such as cancer cells. When incorporated, these compounds tend to cause many mutations in the cells, leading to the death of cancer cells. Unfortunately, other actively dividing cells, such as those in the skin and the lining of the digestive tract, also incorporate 5BU, which leads to unwanted side effects of chemotherapy, such as hair loss and a diminished appetite.

DNA molecules are also sensitive to physical agents such as radiation. In particular, radiation of short wavelength and high energy, known as ionizing radiation, can alter DNA structure. This type of radiation includes X-rays and gamma rays. Ionizing radiation can penetrate deeply into biological materials, where it produces chemically reactive molecules known as free radicals. These molecules alter the structure of DNA in a variety of ways. Exposure to high doses of ionizing radiation results in base deletions, oxidized bases, single nicks in DNA strands, crosslinking, and even chromosomal breaks.

Nonionizing radiation, such as UV light, contains less energy, and so it penetrates only the surface of an organism, such as the skin. Nevertheless, UV light is known to cause DNA mutations. As shown in Figure 18.15, UV light causes the formation of crosslinked **thymine dimers.** Thymine dimers interfere with transcription and DNA replication. This can result in a mutation when that DNA strand is replicated. Plants, in particular, must have effective ways of preventing UV damage because they are exposed to sunlight throughout the day. Sun tanning greatly increases a person's exposure to UV light, raising the potential for thymine dimers and mutation. This explains the higher incidence of skin cancer among people who have been exposed

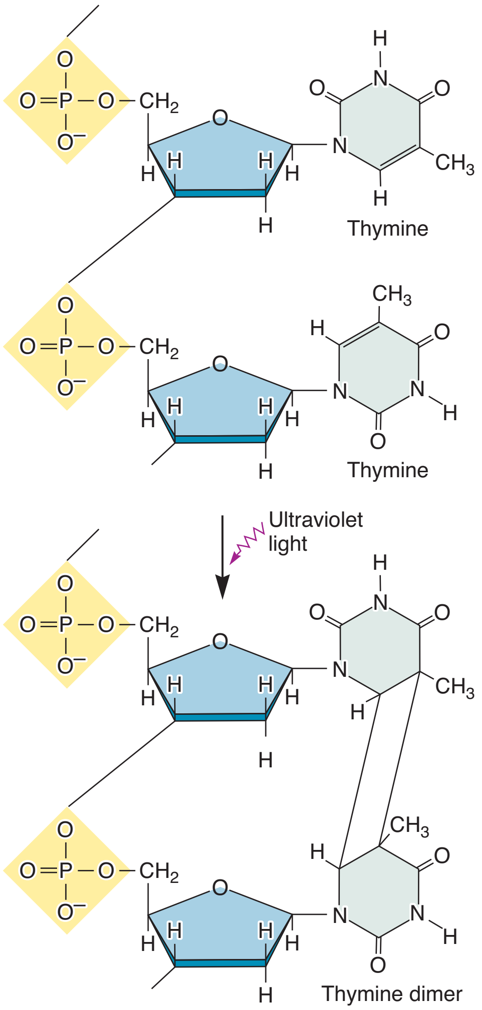

O
O=P—O—CH₂
O⁻
O
H
N
O
O
N
CH₃
H
H
H
H
H
H
Thymine
CH₃
H
O
N
N
H
O
Thymine
Ultraviolet
light
Thymine dimer

FIGURE 18.15 **Formation and structure of a thymine dimer.**

Concept check: *In people, what is a common cause of thymine dimer formation and in what cell type(s) would it be most likely to occur?*

to large amounts of sun during their lifetime. Because of the known link between skin cancer and sun exposure, people now apply sunscreen to their skin to prevent the harmful effects of UV light. Most sunscreens contain organic compounds, such as oxybenzone, which absorb UV light, and/or opaque ingredients, such as zinc oxide, that reflect UV light.

The Mutation Rate Is the Likelihood of a New Mutation

Because mutations occur spontaneously and may be induced by environmental agents, geneticists are greatly interested in learning how prevalent they are. The term **mutation rate** is the likelihood that a gene will be altered by a new mutation. This rate is commonly expressed as the number of new mutations in a given gene per cell generation. The spontaneous mutation rate for a particular gene is typically in the range of 1 in 100,000 to 1 in 1 billion, or 10^{-5} to 10^{-9} per cell generation. In addition, mutations can occur at other sites in a genome, not only in genes. For example, people usually carry about 100 to 200 new mutations in their entire genome compared with their parents. Most of these are single nucleotide changes that do not occur within the coding sequence of genes. With a human genome size of approximately 3,200,000,000 bp, these numbers tell us that a mutation is a relatively infrequent event. However, the mutation rate is not a constant number. The presence of certain environmental agents, such as X-rays, can increase the rate of induced mutations to a much higher value than the spontaneous mutation rate. In addition, mutation rates vary substantially from species to species and even within different strains of the same species. One explanation for this variation is that there are many different causes of mutations (refer back to Table 18.4).

The rate of new mutation is different from the concept of mutation frequency. The **mutation frequency** for a gene is the number of mutant genes divided by the total number of copies of that gene within a population. If 1 million bacteria were plated and 10 were found to be mutant, the mutation frequency would be 1 in 100,000, or 10^{-5}. As we have seen, Luria and Delbrück showed that among the bacteria in the 20 tubes, the timing of mutations influenced the mutation frequency within any particular tube. Some tubes had a high frequency of mutation, but others did not. Therefore, the mutation frequency depends not only on the mutation rate but also on the timing of mutation, and on the likelihood that a mutation will be passed to future generations. The mutation frequency is an important genetic concept, particularly in the field of population genetics. As we will see in Chapter 25, mutation frequencies may rise above the mutation rate due to evolutionary factors such as natural selection and genetic drift.

Testing Methods Can Determine If an Agent Is a Mutagen

To determine if an agent is mutagenic, researchers use testing methods that monitor whether or not an agent increases the mutation rate. Many different kinds of tests have been used to evaluate mutagenicity. One common method is the **Ames test,** which was developed by Bruce Ames. This test uses strains of a bacterium, *Salmonella typhimurium*, that cannot synthesize the amino acid histidine. These strains contain a point mutation within a gene that encodes an enzyme required for histidine biosynthesis. The mutation renders the enzyme inactive. Therefore, the bacteria cannot grow on petri plates unless histidine has been added to the growth medium. However, a second mutation—a reversion—may occur that restores the ability to synthesize histidine. In other words, a second mutation can cause a reversion back to the wild-type condition. The Ames test monitors the rate at which this second mutation occurs, thereby indicating whether an agent increases the mutation rate above the spontaneous rate.

Figure 18.16 outlines the steps in the Ames test. The suspected mutagen is mixed with a rat liver extract and a strain of *Salmonella* that cannot synthesize histidine. A mutagen may require activation by cellular enzymes, which are provided by the rat liver extract. This step improves the ability of the test to identify agents that may cause mutation in mammals. After the incubation period, a large number of bacteria are then plated on a growth medium that does not contain histidine. The *Salmonella* strain is not expected to grow on these plates. However, if a mutation has occurred that allows the strain to synthesize histidine, it can grow on these plates

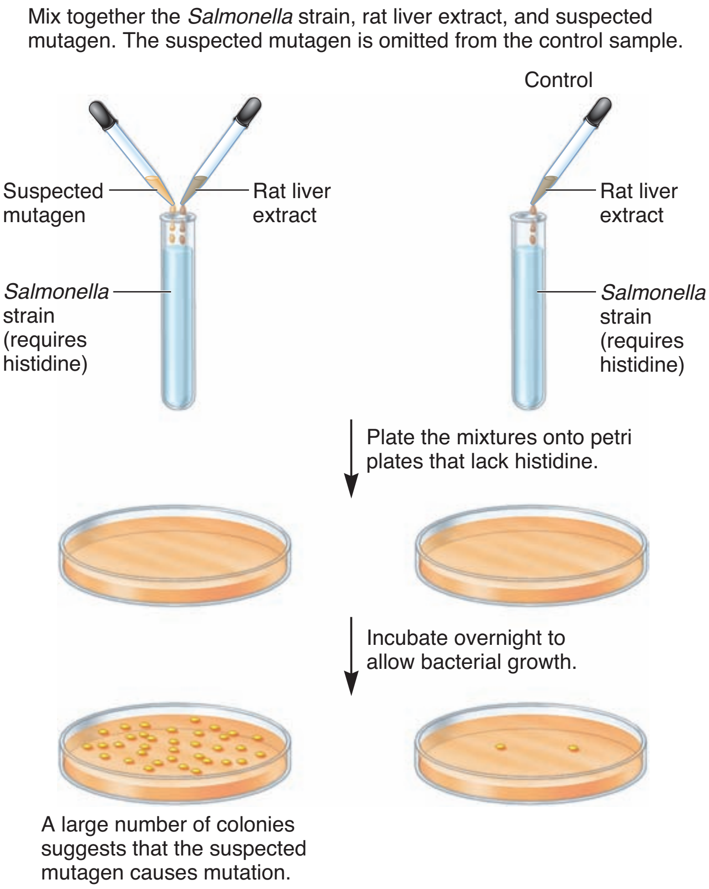

Mix together the *Salmonella* strain, rat liver extract, and suspected mutagen. The suspected mutagen is omitted from the control sample.
Control
Suspected mutagen
Rat liver extract
Rat liver extract
Salmonella strain (requires histidine)
Salmonella strain (requires histidine)
Plate the mixtures onto petri plates that lack histidine.
Incubate overnight to allow bacterial growth.
A large number of colonies suggests that the suspected mutagen causes mutation.

FIGURE 18.16 **The Ames test for mutagenicity.**

Concept check: What is the purpose of the rat liver extract in this procedure?

to form a visible bacterial colony. To estimate the mutation rate, the colonies that grow on the media are counted and compared with the total number of bacterial cells that were originally streaked on the plate. For example, if 10,000,000 bacteria were plated and 10 growing colonies were observed, the rate of mutation is 10 out of 10,000,000; this equals 1 in 10^6, or simply 10^{-6}. As a control, bacteria that have not been exposed to the mutagen are also tested, because a low level of spontaneous mutations is expected to occur.

How do we judge if an agent is a mutagen? Researchers compare the mutation rate in the presence and absence of the suspected mutagen. The experimental approach shown in Figure 18.16 is conducted several times. If statistics reveal that the mutation rate in the experimental and control samples are significantly different, researchers may tentatively conclude that the agent may be a mutagen. Many studies have been conducted in which researchers used the Ames test to compare the urine from cigarette smokers to that from nonsmokers. This research has shown that the urine from smokers contains much higher levels of mutagens.

As discussed in Solved Problem S3 at the end of this chapter, several commercially available *his*$^-$ strains of *Salmonella* carry different kinds of mutations within the coding sequence. This makes it possible to determine if a mutagen causes transitions, transversions, or frameshift mutations.

18.4 REVIEWING THE KEY CONCEPTS

- A mutagen is an agent that can cause a mutation. Researchers have identified many different mutagens that change DNA structure in a variety of ways (see Table 18.6, Figures 18.13–18.15).
- Mutation rate is the likelihood that a new mutation will occur. Mutation frequency is the number of mutant genes divided by the total number of genes in a population.
- The Ames test is used to determine if an agent is a mutagen (see Figure 18.16).

18.4 COMPREHENSION QUESTIONS

1. Nitrous acid replaces amino groups with keto groups, a process called
   a. alkylation.
   b. deamination.
   c. depurination.
   d. crosslinking.
2. A mutagen that is a base analog is
   a. ethyl methanesulfonate (EMS).
   b. 5-bromouracil.
   c. UV light.
   d. proflavin.
3. In the fluctuation test of Luria and Delbrück described in Section 18.2, the number of colonies on individual plates was a measure of
   a. the mutation rate.
   b. the mutation frequency.
   c. both a and b.
   d. none of the above.
4. In the Ames test, a ____________ number of colonies is observed if a substance ________ a mutagen, compared with the control that is not exposed to the suspected mutagen.
   a. significantly higher, is
   b. significantly higher, is not
   c. significantly lower, is
   d. significantly lower, is not

18.5 DNA REPAIR

Because most mutations are deleterious, DNA repair systems are vital to the survival of all organisms. If DNA repair systems did not exist, spontaneous and environmentally induced mutations would be so prevalent that few species, if any, would survive. The necessity of DNA repair systems becomes evident when they are missing. Bacteria contain several different DNA repair systems. Yet, when even a single system is absent, the bacteria have a much higher rate of mutation. In fact, the rate of mutation is so high that these bacterial strains are sometimes called mutator strains. Likewise, in humans, an individual who is defective in only a single DNA repair system may manifest various disease symptoms, including a higher risk of skin cancer. This increased risk is due to the inability to repair UV-induced mutations.

Living cells produce several DNA repair systems that can fix different types of DNA alterations (**Table 18.7**). Each repair system is composed of one or more proteins that play specific roles in the repair mechanism. In most cases, DNA repair is a multistep process. First, one or more proteins in the DNA repair system detect an irregularity in DNA structure. Next, the abnormality is removed by the action of DNA repair enzymes. Finally, normal DNA is synthesized via DNA replication enzymes. In this section, we will examine several different repair systems that have been characterized in bacteria, yeast, mammals, and plants. Their diverse ways of repairing DNA underscore the extreme necessity for the structure of DNA to be maintained properly.

Damaged Bases Can Be Directly Repaired

In a few cases, the covalent modification of nucleotides by mutagens can be reversed by specific cellular enzymes. As discussed earlier in this chapter, UV light causes the formation of thymine dimers. Bacteria, fungi, most plants, and some animals produce an enzyme called **photolyase** that recognizes thymine dimers and splits them, which returns the DNA to its original condition (**Figure 18.17a**). Photolyase contains two light-sensitive cofactors. The repair mechanism itself requires light and is known as **photoreactivation.** This process directly restores the structure of DNA. Because plants are exposed to sunlight throughout the day, photolyase is a critical DNA repair enzyme for many plant species.

A protein known as **alkyltransferase** can remove methyl or ethyl groups from guanine bases that have been mutagenized by alkylating agents such as nitrogen mustard and EMS. This protein is called alkyltransferase because it transfers the methyl

TABLE 18.7

Common Types of DNA Repair Systems

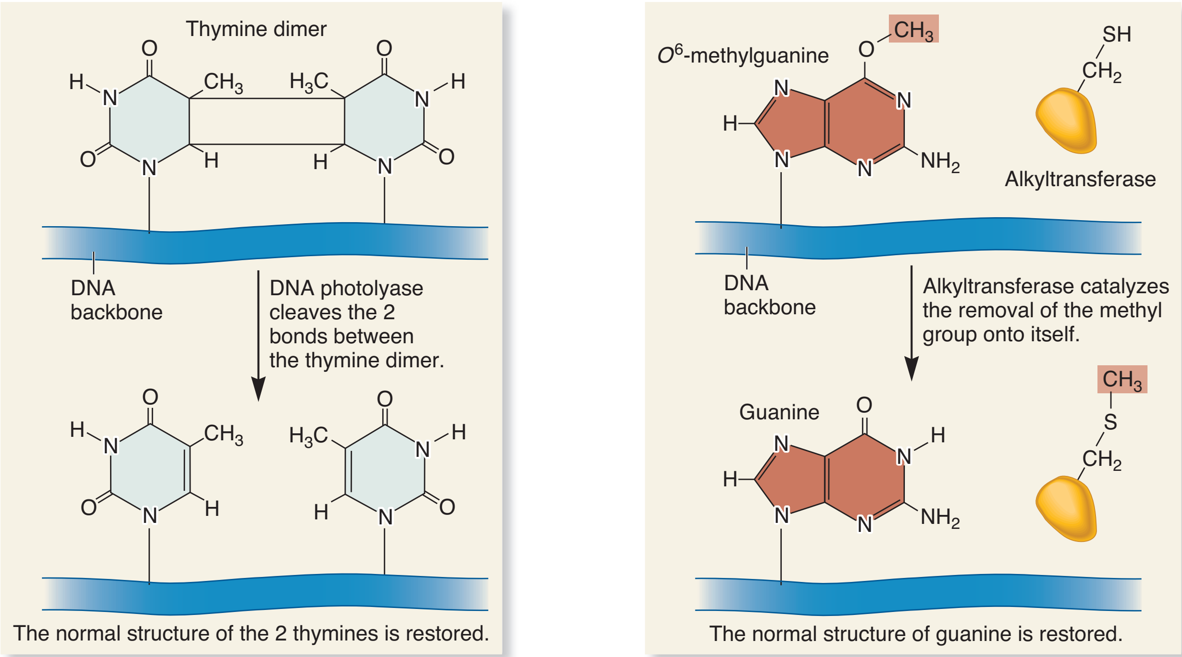
System	Description
Direct repair	An enzyme recognizes an incorrect alteration in DNA structure and directly converts it back to a correct structure.
Base excision repair and nucleotide excision repair	An abnormal base or nucleotide is first recognized and removed from the DNA, and a segment of DNA in this region is excised, and then the complementary DNA strand is used as a template to synthesize a normal DNA strand.
Mismatch repair	Similar to excision repair except that the DNA defect is a base pair mismatch in the DNA, not an abnormal nucleotide. The mismatch is recognized, and a segment of DNA in this region is removed. The parental strand is used to synthesize a normal daughter strand of DNA.
Homologous recombination repair	Occurs at double-strand breaks or when DNA damage causes a gap in synthesis during DNA replication. The strands of a normal sister chromatid are used to repair a damaged sister chromatid.
Nonhomologous end joining	Occurs at double-strand breaks. The broken ends are recognized by proteins that keep the ends together; the broken ends are eventually rejoined.

or ethyl group from the base to a cysteine side chain within the alkyltransferase protein (**Figure 18.17b**). Surprisingly, this permanently inactivates alkyltransferase, which means it can be used only once!

Base Excision Repair Removes a Damaged Base

A second type of repair system, called **base excision repair (BER),** involves the function of a category of enzymes known as **DNA *N*-glycosylases.** This type of enzyme can recognize an abnormal base and cleave the bond between it and the sugar in the DNA backbone, creating an apurinic or apyrimidinic site (**Figure 18.18**). Living organisms produce multiple types of DNA *N*-glycosylases, each recognizing particular types of abnormal base structures. Depending on the DNA *N*-glycosylase, this repair system can eliminate abnormal bases such as uracil, 3-methyladenine, 7-methylguanine, and thymine dimers. BER is particularly important for the repair of oxidative DNA damage.

Figure 18.18 illustrates the general steps involved in DNA repair via *N*-glycosylase. In this example, the DNA contains a uracil in its sequence. This could have happened spontaneously or by the action of a chemical mutagen. *N*-glycosylase recognizes a uracil within the DNA and cleaves (nicks) the bond between the sugar and base. This releases the uracil base and leaves behind an apyrimidinic site. This abnormality is recognized by a second enzyme, **AP endonuclease,** which makes a cut on the 5′ side.


Thymine dimer
O
O
H
N
CH3
H3C
N
H
O
N
H
H
N
O
DNA backbone
DNA photolyase cleaves the 2 bonds between the thymine dimer.
The normal structure of the 2 thymines is restored.
O6-methylguanine
CH3
O
SH
CH2
N
N
H
N
N
NH2
Alkyltransferase
DNA backbone
Alkyltransferase catalyzes the removal of the methyl group onto itself.
Guanine
O
CH3
S
CH2
N
N
H
NH2
The normal structure of guanine is restored.

(a) Direct repair of a thymine dimer

(b) Direct repair of a methylated base

ONLINE ANIMATION

FIGURE 18.17 **Direct repair of damaged bases in DNA.** **(a)** The repair of thymine dimers by photolyase. **(b)** The repair of methylguanine by the transfer of the methyl group to alkyltransferase.

Concept check: *Which of these repair systems is particularly valuable to plants?*

Following this cut by AP endonuclease, one of three things can happen. In some species such as *E. coli*, DNA polymerase I, which has a 5′ to 3′ exonuclease activity, removes a DNA segment containing the abnormal region and, at the same time, replaces it with normal nucleotides. This process is called **nick translation** (although DNA replication, not mRNA translation, actually occurs). Alternatively, in eukaryotic species such as humans, the DNA is repaired in two possible ways. DNA polymerase β has the enzymatic ability to remove a site that is missing a base, and then insert a nucleotide with the correct base in its place. Another possibility is that DNA polymerase δ or ε can synthesize a short segment of DNA, which generates a flap. The flap is then removed by flap endonuclease, which is described in Chapter 13. In all three cases, the final step is carried out by DNA ligase, which closes a gap in the DNA backbone.

Nucleotide Excision Repair Systems Remove Segments of Damaged DNA

An important general process for DNA repair is the **nucleotide excision repair (NER)** system. This type of system can repair many different types of DNA damage, including thymine dimers, chemically modified bases, missing bases, and certain types of crosslinks. In NER, several nucleotides in the damaged strand are removed from the DNA, and the intact strand is used as a template for resynthesis of a normal complementary strand. NER

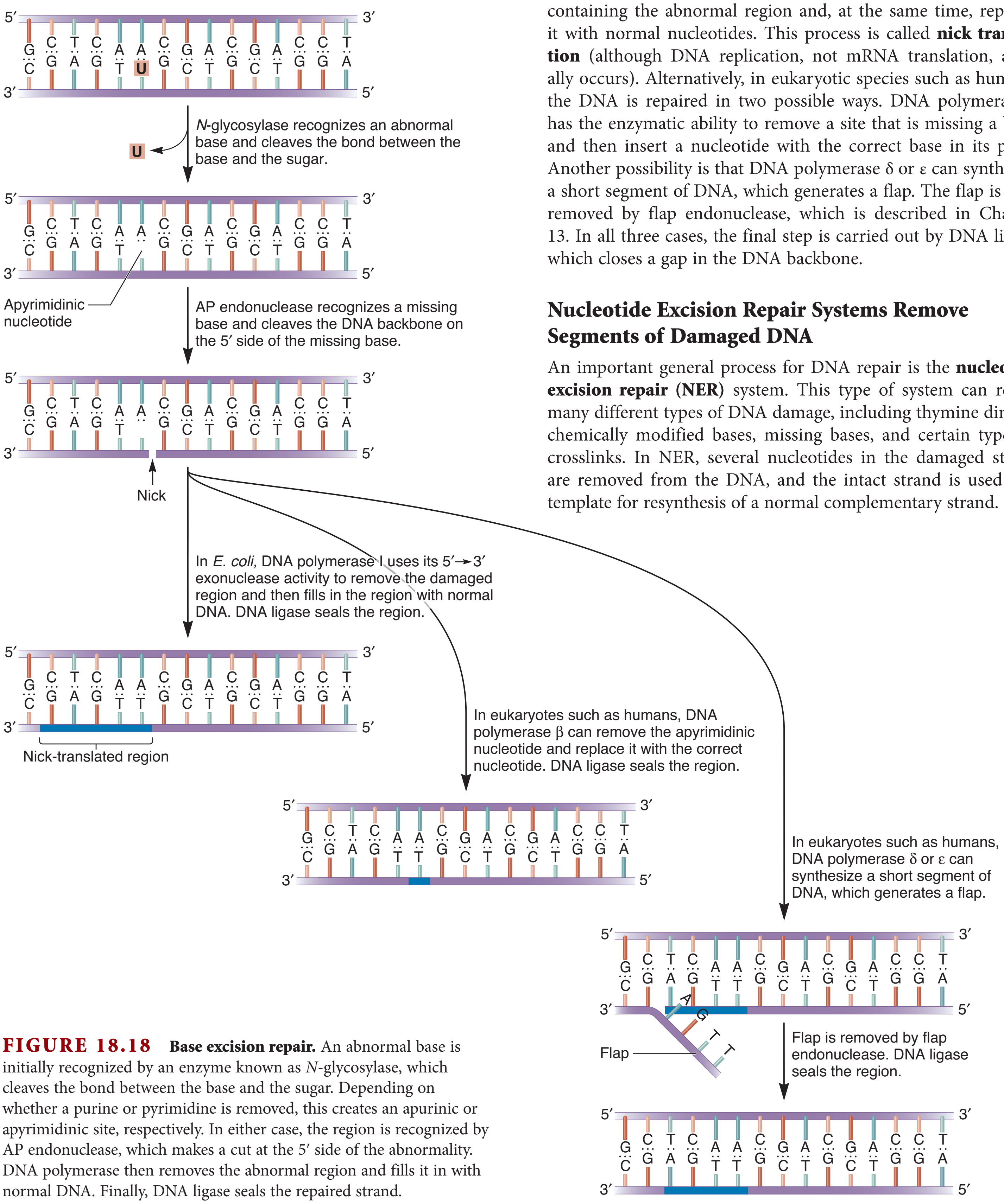

5′
3′
U
N-glycosylase recognizes an abnormal base and cleaves the bond between the base and the sugar.
Apyrimidinic nucleotide
AP endonuclease recognizes a missing base and cleaves the DNA backbone on the 5′ side of the missing base.
Nick
In *E. coli,* DNA polymerase I uses its 5′→3′ exonuclease activity to remove the damaged region and then fills in the region with normal DNA. DNA ligase seals the region.
Nick-translated region
In eukaryotes such as humans, DNA polymerase β can remove the apyrimidinic nucleotide and replace it with the correct nucleotide. DNA ligase seals the region.
In eukaryotes such as humans, DNA polymerase δ or ε can synthesize a short segment of DNA, which generates a flap.
Flap
Flap is removed by flap endonuclease. DNA ligase seals the region.

FIGURE 18.18 **Base excision repair.** An abnormal base is initially recognized by an enzyme known as *N*-glycosylase, which cleaves the bond between the base and the sugar. Depending on whether a purine or pyrimidine is removed, this creates an apurinic or apyrimidinic site, respectively. In either case, the region is recognized by AP endonuclease, which makes a cut at the 5′ side of the abnormality. DNA polymerase then removes the abnormal region and fills it in with normal DNA. Finally, DNA ligase seals the repaired strand.

is found in all eukaryotes and prokaryotes, although its molecular mechanism is better understood in prokaryotic species.

In *E. coli*, the NER system requires four key proteins, designated UvrA, UvrB, UvrC, and UvrD, plus the help of DNA polymerase and DNA ligase. UvrA, B, C, and D recognize and remove a short segment of a damaged DNA strand. They are named Uvr because they are involved in Ultraviolet light repair of thymine dimers, although the UvrA–D proteins are also important in repairing chemically damaged DNA.

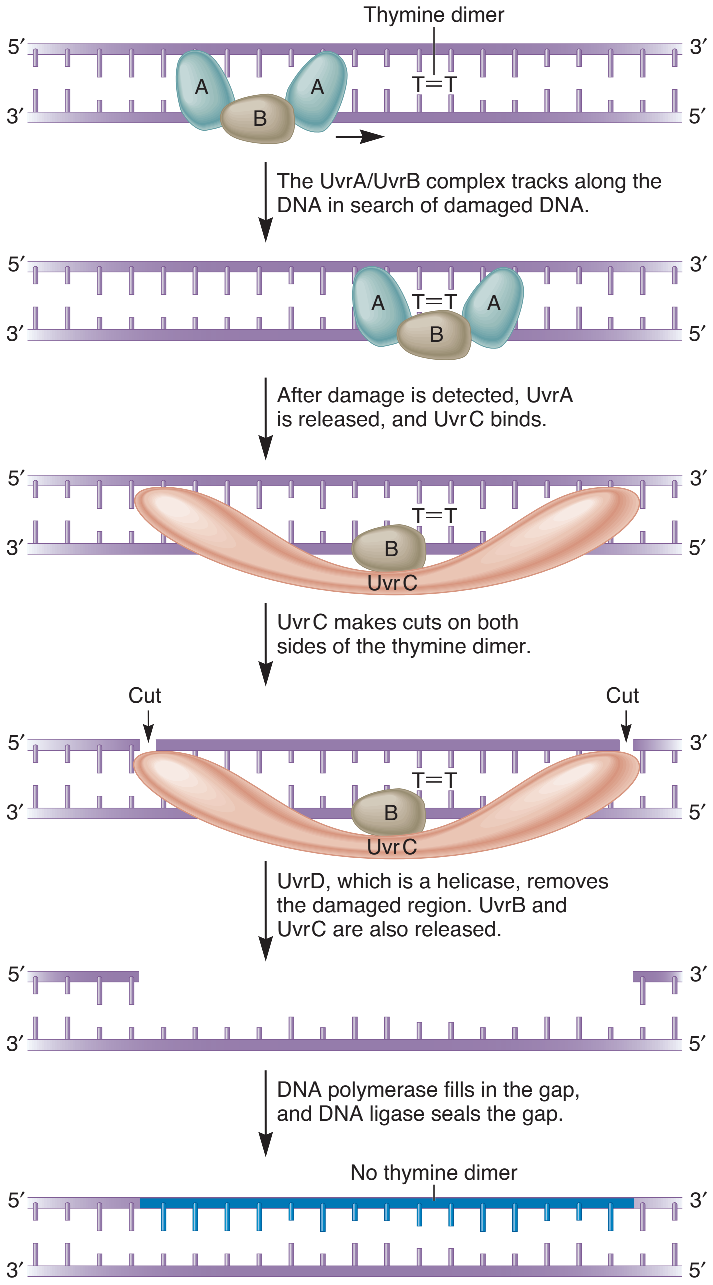
Figure 18.19 outlines the steps involved in the *E. coli* NER system. A protein complex consisting of two UvrA molecules and one UvrB molecule tracks along the DNA in search of damaged DNA. Such DNA has a distorted double helix, which is sensed by the UvrA/UvrB complex. When a damaged segment is identified, the two UvrA proteins are released, and UvrC binds to the site. The UvrC protein makes cuts in the damaged strand on both sides of the damaged site. Typically, the damaged strand is cut eight nucleotides from the 5′ end of the damage and four to five nucleotides away from the 3′ end. After this process, UvrD, which is a helicase, recognizes the region and separates the two strands of DNA. This releases a short DNA segment that contains the damaged region, and UvrB and UvrC are also released. Following the excision of the damaged DNA, DNA polymerase fills in the gap, using the undamaged strand as a template. Finally, DNA ligase makes the final covalent connection between the newly made DNA and the original DNA strand.

In eukaryotes, NER systems are thought to operate similarly to bacterial NER systems, though more proteins are involved. Several human diseases are due to inherited defects in genes involved in NER. These include xeroderma pigmentosum (XP), Cockayne syndrome (CS), and PIBIDS. (PIBIDS is an acronym for a syndrome with symptoms that include photosensitivity; ichthyosis, a skin abnormality; brittle hair; impaired intelligence; decreased fertility; and short stature.) A common characteristic in all three syndromes is an increased sensitivity to sunlight because of an inability to repair UV-induced lesions. **Figure 18.20** is a photograph of an individual with XP. Such individuals have pigmentation abnormalities, many premalignant lesions, and a high predisposition to skin cancer. They may also develop early degeneration of the nervous system.

Genetic analyses of patients with XP, CS, and PIBIDS have revealed that these syndromes result from defects in a variety of different genes that encode NER proteins. For example, XP can be caused by defects in any of seven different NER genes. In all cases, individuals have a defective NER pathway. In recent years, several human NER genes have been successfully cloned and sequenced. Although more research is needed to completely understand the mechanism of DNA repair, the identification of NER genes has helped unravel the complexities of NER pathways in human cells.

Mismatch Repair Systems Recognize and Correct a Base Pair Mismatch

Thus far, we have considered several DNA repair systems that recognize abnormal nucleotide structures within DNA, including


Thymine dimer
5′
3′
A
A
B
T=T
3′
5′
The UvrA/UvrB complex tracks along the DNA in search of damaged DNA.
5′
3′
A
T=T
A
B
3′
5′
After damage is detected, UvrA is released, and UvrC binds.
5′
3′
T=T
B
UvrC
3′
5′
UvrC makes cuts on both sides of the thymine dimer.
Cut
Cut
5′
3′
T=T
B
UvrC
3′
5′
UvrD, which is a helicase, removes the damaged region. UvrB and UvrC are also released.
5′
3′
3′
5′
DNA polymerase fills in the gap, and DNA ligase seals the gap.
No thymine dimer
5′
3′
3′
5′

ONLINE ANIMATION **FIGURE 18.19 Nucleotide excision repair in *E. coli*.** The UvrA/UvrB complex scans along the DNA. When the complex encounters a thymine dimer in the DNA, the UvrA dimer is released. UvrC binds to UvrB. UvrC makes two cuts in the damaged DNA strand. One cut is made eight nucleotides 5′ to the site, the other approximately five nucleotides 3′ from the site. UvrD (not shown in the figure) binds to the site. UvrD is a helicase that causes the damaged strand to be removed. During this process, UvrB and UvrC are released. DNA polymerase resynthesizes a complementary strand. DNA ligase makes the final connection.

Concept check: Explain why cuts are made on both sides of the DNA damage.

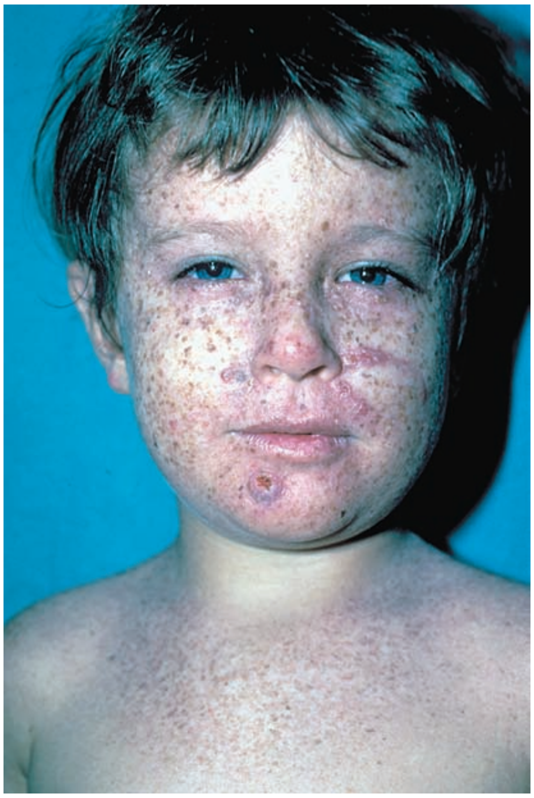


FIGURE 18.20 **An individual affected with xeroderma pigmentosum.**

Genes→Traits This disease involves a defect in genes that are involved with nucleotide excision repair (NER). Xeroderma pigmentosum can be caused by defects in seven different NER genes. Affected individuals have an increased sensitivity to sunlight because of an inability to repair UV-induced DNA lesions. In addition, they may also have pigmentation abnormalities, many premalignant lesions, and a high predisposition to skin cancer.

thymine dimers, alkylated bases, and the presence of uracil in the DNA. Another type of abnormality that should not occur in DNA is a **base pair mismatch.** The structure of the DNA double helix obeys the AT/GC rule of base pairing. During the normal course of DNA replication, however, an incorrect nucleotide may be added to the growing strand by mistake. This produces a mismatch between a nucleotide in the parental and the newly made strand. Various DNA repair mechanisms can recognize and remove this mismatch. For example, as described in Chapter 13, DNA polymerase has a 3′ to 5′ proofreading ability that detects mismatches and removes them. However, if this proofreading ability fails, cells have additional DNA repair systems that detect base mismatches and fix them. An interesting DNA repair system that exists in all species is the **mismatch repair system.**

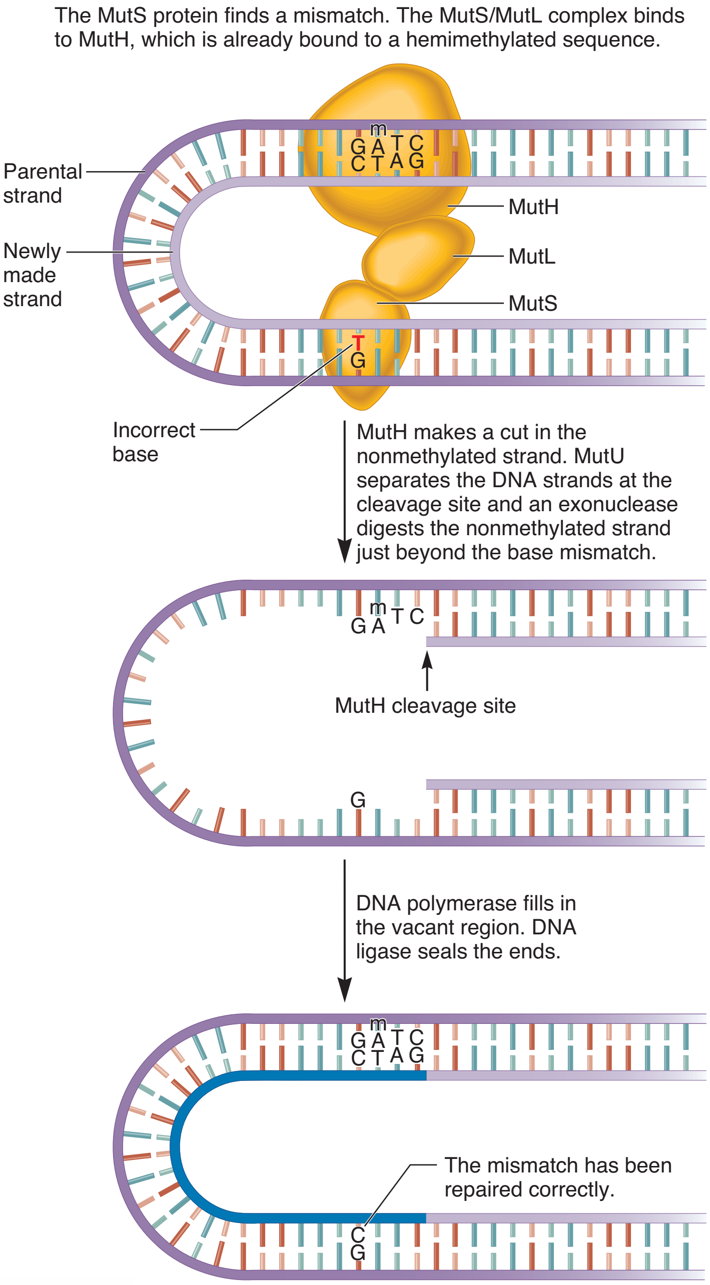
In the case of a base mismatch, how does a DNA repair system determine which base to remove? If the mismatch is due to an error in DNA replication, the newly made daughter strand contains the incorrect base, whereas the parental strand is normal. Therefore, an important aspect of mismatch repair is that it specifically repairs the newly made strand rather than the parental template strand. Prior to DNA replication, the parental DNA has already been methylated. Immediately after DNA replication, some time must pass before a newly made strand is methylated. Therefore, newly replicated DNA is hemimethylated—only the parental DNA strand is methylated. Hemimethylation provides a way for a DNA repair system to distinguish between the parental DNA strand and the daughter strand.

The molecular mechanism of mismatch repair has been studied extensively in *E. coli*. As shown in **Figure 18.21**, three proteins, designated MutS, MutL, and MutH, detect the mismatch and direct the removal of the mismatched base from the newly made strand. These proteins are named Mut because their absence leads to a much higher mutation rate than occurs in normal strains of *E. coli*. The role of MutS is to locate mismatches. Once a mismatch is detected, MutS forms a complex with MutL. MutL acts as a linker that binds to MutH, forming a loop in the DNA. This stimulates MutH, which is bound to a hemimethylated site, to make a cut in the newly made, nonmethylated DNA strand. After the strand is cut, MutU, which functions as helicase, separates the strands, and an exonuclease then digests the nonmethylated DNA strand in the direction of the mismatch and proceeds beyond the mismatch site. This leaves a gap in the daughter strand that is repaired by DNA polymerase and DNA ligase. The net result is that the mismatch has been corrected by removing the incorrect region in the daughter strand and then resynthesizing the correct sequence using the parental DNA as a template.

Eukaryotic species also have homologs to MutS and MutL, along with many other proteins that are needed for mismatch repair. However, a MutH homolog has not been identified in eukaryotes, and the mechanism by which the eukaryotic mismatch repair system distinguishes between the parental and daughter strands is not well understood. As with defects in nucleotide excision repair systems, mutations in the human mismatch repair system are associated with particular types of cancer. For example, mutations in two human mismatch repair genes, *hMSH2* and *hMLH1*, play a role in the development of a type of colon cancer known as hereditary nonpolyposis colorectal cancer.

Double-Strand Breaks Can Be Repaired by Homologous Recombination Repair and by Nonhomologous End Joining

Of the many types of DNA damage that can occur within living cells, the breakage of chromosomes—called a DNA double-strand break (DSB)—is perhaps the most dangerous. DSBs can be caused by ionizing radiation (X-rays or gamma rays), chemical mutagens, and certain drugs used for chemotherapy. In addition, reactive oxygen species that are the by-products of aerobic metabolism cause double-strand breaks. Surprisingly, researchers estimate that naturally occurring double-strand breaks in a typical human cell occur at a rate of between 10 and 100 breaks per cell per day! Such breaks are harmful in a variety of ways. First, DSBs can result in chromosomal rearrangements such as inversions and translocations (refer back to Figure 8.2). In addition, DSBs can lead to terminal or interstitial deletions (refer back to Figure 8.3). Such genetic changes have the potential to result in detrimental phenotypic effects.


The MutS protein finds a mismatch. The MutS/MutL complex binds to MutH, which is already bound to a hemimethylated sequence.
Parental strand
Newly made strand
m
GATC
CTAG
MutH
MutL
MutS
T
G
Incorrect base
MutH makes a cut in the nonmethylated strand. MutU separates the DNA strands at the cleavage site and an exonuclease digests the nonmethylated strand just beyond the base mismatch.
m
GATC
MutH cleavage site
G
DNA polymerase fills in the vacant region. DNA ligase seals the ends.
m
GATC
CTAG
The mismatch has been repaired correctly.
C
G

ONLINE ANIMATION

FIGURE 18.21 **Mismatch repair in *E. coli.*** MutS slides along the DNA and recognizes base mismatches in the double helix. MutL binds to MutS and acts as a linker between MutS and MutH. The DNA must loop for this interaction to occur. The role of MutH is to identify the methylated strand of DNA, which is the nonmutated parental strand. MutH can be bound to a methylated site on either side of the base mismatch. In this example, the methylated adenine is designated with an m. MutH makes a cut in the nonmethylated strand. MutU (not shown in the figure) begins to separate the strands at the cleavage site, and then an exonuclease digests the nonmethylated strand until it passes the MutS/MutL region. This leaves a gap, which is filled in by DNA polymerase and sealed by DNA ligase.

Concept check: Which of these three proteins is responsible for ensuring that the mismatched base in the newly made daughter strand is the one that is removed?

How are DSBs repaired? The two main mechanisms are **homologous recombination repair (HRR)** and **nonhomologous end joining (NHEJ).** HRR occurs when homologous DNA strands, usually from a sister chromatid, are used to repair a DSB in the other sister chromatid (**Figure 18.22**). First, the DSB is processed by the short digestion of DNA strands at the

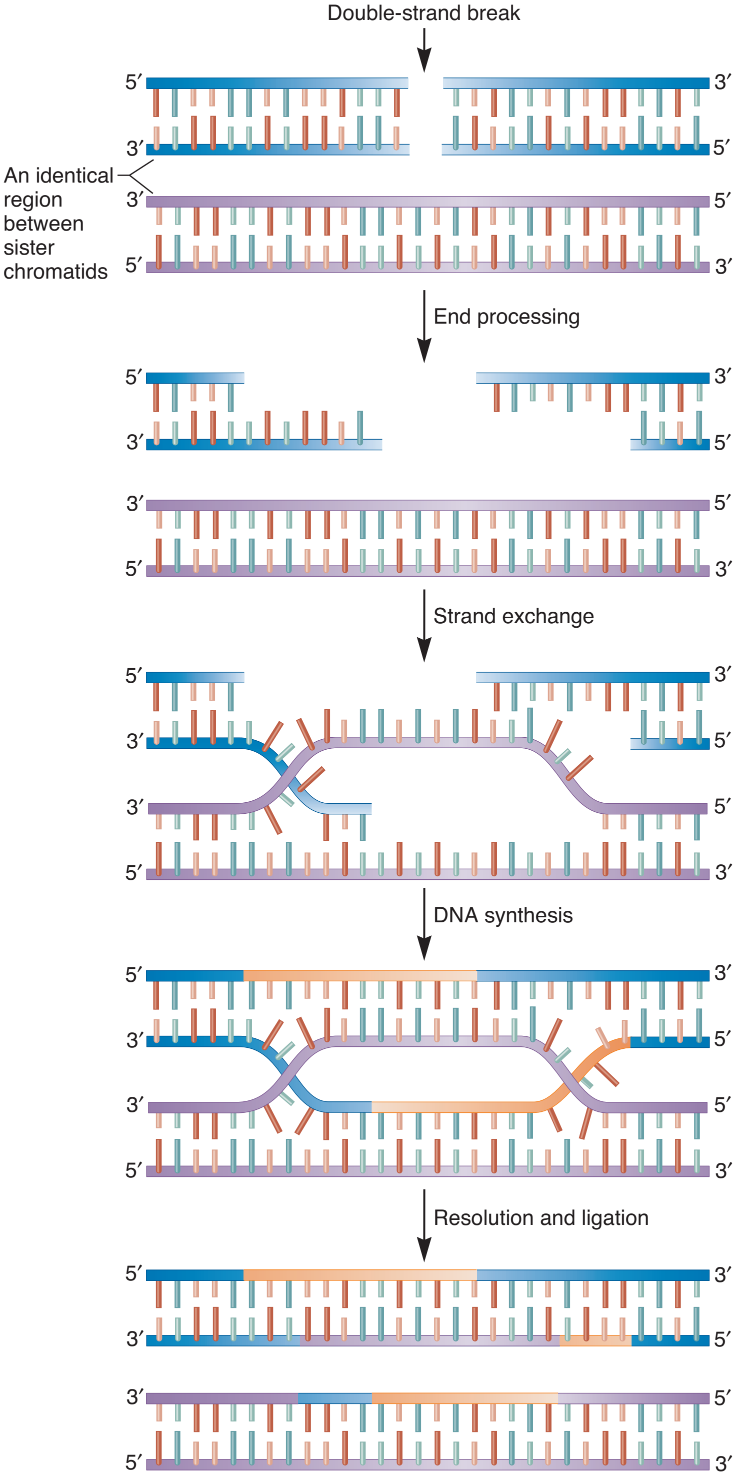

Double-strand break
5′
3′
3′
5′
An identical region between sister chromatids
3′
5′
5′
3′
End processing
5′
3′
3′
5′
3′
5′
5′
3′
Strand exchange
5′
3′
3′
5′
3′
5′
5′
3′
DNA synthesis
5′
3′
3′
5′
3′
5′
5′
3′
Resolution and ligation
5′
3′
3′
5′
3′
5′
5′
3′

FIGURE 18.22 **DNA repair of a double-strand break (DSBs) via homologous recombination repair.**

break site. This processing event is followed by the exchange of DNA strands between the broken and unbroken sister chromatids. The unbroken strands are then used as templates to synthesize DNA in the region where the break occurred. Finally, the crisscrossed strands are resolved, which means they are broken and then rejoined in a way that produces separate chromatids. Because sister chromatids are genetically identical, an advantage is that homologous recombination repair can be an error-free mechanism for repairing a DSB. A disadvantage is that sister chromatids are available only during the S and G_2 phases of the cell cycle in eukaryotes or following DNA replication in bacteria. Although sister chromatids are strongly preferred, HRR may also occur between homologous regions that are not identical. Therefore, HRR may occasionally happen when sister chromatids are unavailable. The proteins involved in homologous recombination repair are described in Chapter 13 (see Table 13.5).

During nonhomologous end joining, the two broken ends of DNA are simply pieced back together (**Figure 18.23**). This mechanism requires the participation of several proteins that play key roles in the process. First, the DSB is recognized by end-binding proteins. These proteins then recognize additional proteins that form a cross-bridge that prevents the two ends from drifting apart. Next, additional proteins are recruited to the region that may process the ends of the broken chromosome by digesting particular DNA strands. This processing may result in the deletion of a small amount of genetic material from the region. Finally, any gaps are filled in via DNA polymerase, and the DNA ends are ligated together. One advantage of NHEJ is that it doesn't involve the participation of a sister chromatid, so it can occur at any stage of the cell cycle. However, a disadvantage is that NHEJ can result in small deletions in the region that has been repaired.

Actively Transcribed DNA Is Repaired More Efficiently Than Nontranscribed DNA

Not all DNA is repaired at the same rate. In the 1980s, Philip Hanawalt and colleagues conducted experiments showing that actively transcribed genes in eukaryotes and prokaryotes are more efficiently repaired following radiation damage than is nontranscribed DNA. This phenomenon is called **transcription-coupled DNA repair.** The targeting of DNA repair enzymes to actively transcribing genes may have several biological advantages. First, active genes are loosely packed and may be more vulnerable to DNA damage. Likewise, the process of transcription itself may make the DNA more susceptible to agents that cause damage. In addition, DNA regions that contain actively transcribed genes are more likely to be important for survival than nontranscribed regions. This may be particularly true in differentiated cells, such as nerve cells, which no longer divide but actively transcribe genes for many years. In nondividing cells, gene transcription rather than DNA replication is of utmost importance.

How are actively transcribed regions targeted by DNA repair systems? In *E. coli*, a protein known as **transcription-repair coupling factor (TRCF)** is responsible for targeting the nucleotide excision repair (NER) system to actively transcribed genes with damaged DNA. **Figure 18.24** illustrates a case where RNA polymerase had been transcribing a gene until it encountered a thymine dimer and became stalled. TRCF is a helicase that displaces the stalled RNA polymerase from the damaged template strand. TRCF also has a binding site for UvrA. Therefore, after RNA polymerase has been removed, TRCF recruits the NER system to the damaged site by recognizing the UvrA/UvrB complex. After the UvrA/UvrB complex has bound to this site, TRCF is released from the damaged site. At this stage, the UvrA/UvrB complex can begin NER, as described earlier in Figure 18.19.

In eukaryotes, transcription-coupled DNA repair also occurs via NER and other DNA repair systems. However, the mechanism that couples transcription and DNA repair is not completely understood. Several different proteins are thought to act as TRCFs. Some of these have been identified in people with high rates of mutation. As mentioned previously, Cockayne

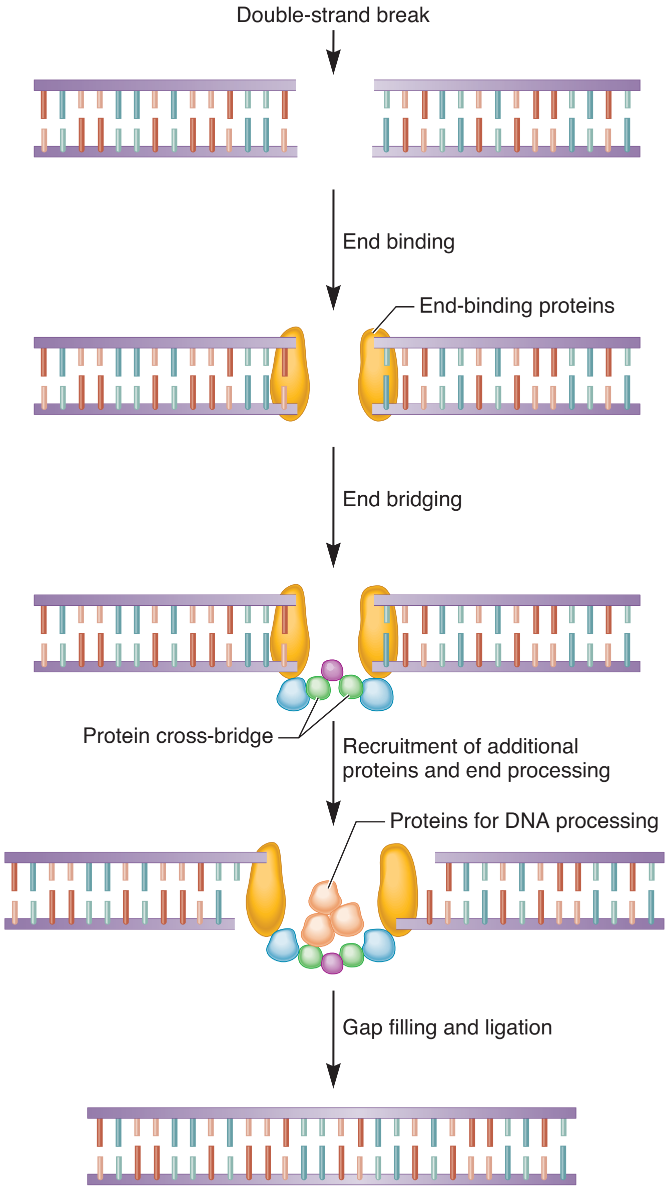


FIGURE 18.23 **DNA repair of a double-strand break via nonhomologous end joining.**

Concept check: *What is an advantage and disadvantage of this repair system?*

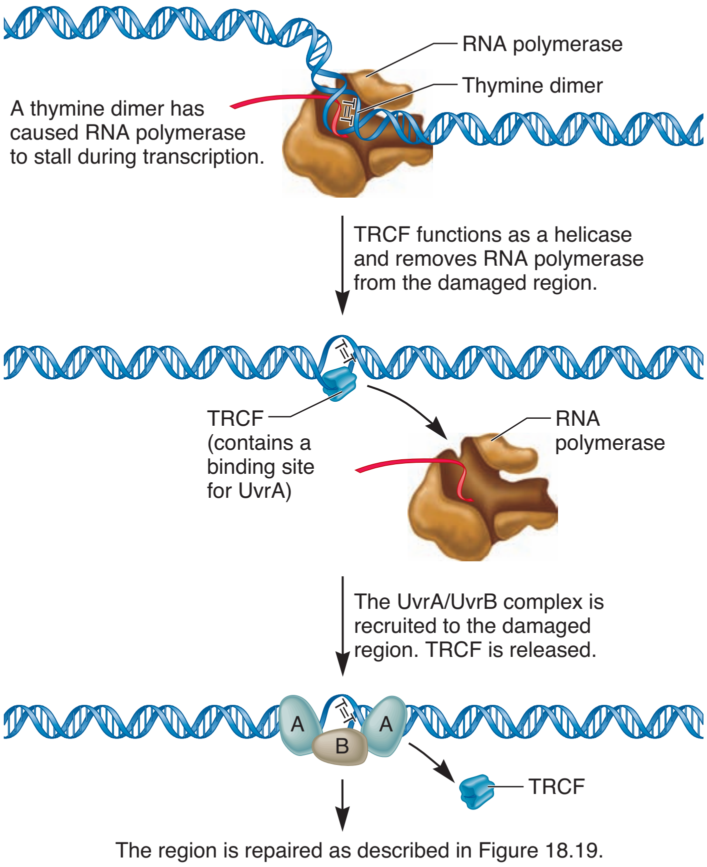

RNA polymerase
Thymine dimer
A thymine dimer has caused RNA polymerase to stall during transcription.
TRCF functions as a helicase and removes RNA polymerase from the damaged region.
TRCF (contains a binding site for UvrA)
RNA polymerase
The UvrA/UvrB complex is recruited to the damaged region. TRCF is released.
A
B
A
TRCF
The region is repaired as described in Figure 18.19.

FIGURE 18.24 **Targeting of the nucleotide excision repair system to an actively transcribing gene containing a DNA alteration.** When RNA polymerase encounters damaged DNA, it stalls over the damaged site. TRCF, which is a helicase, recognizes a stalled RNA polymerase and removes it from the damaged site. TRCF has a binding site for UvrA and thereby recruits the UvrA/UvrB complex to the region. This initiates the process of nucleotide excision repair.

Concept check: What are two advantages of specifically targeting a DNA repair system to actively transcribing genes?

syndrome (CS) involves defects in genes that play a role in NER. Two genes that are defective in CS, termed *CS-A* and *CS-B*, encode proteins that function as TRCFs. In addition, certain general transcription factors, such as TFIIH, play a role in both transcription and NER.

Damaged DNA May Be Replicated by Translesion DNA Polymerases

Despite the efficient action of numerous repair systems that remove lesions in DNA in an error-free manner, it is inevitable that some lesions may escape these repair mechanisms. Such lesions may be present when DNA is being replicated. If so, replicative DNA polymerases, such as polIII in *E. coli*, are highly sensitive to geometric distortions in DNA and are unable to replicate through DNA lesions. During the past decade, researchers have discovered that cells are equipped with specialized DNA polymerases that assist the replicative DNA polymerase during the process of **translesion synthesis (TLS)**—the synthesis of DNA over a template strand that harbors some type of DNA damage. These lesion-replicating polymerases, which are also described in Chapter 13 (see Table 13.3), contain an active site with a loose, flexible pocket that can accommodate aberrant structures in the template strand. When a replicative DNA polymerase encounters a damaged region, it is swapped with a lesion-replicating polymerase.

A negative consequence of translesion synthesis is low fidelity. Due to their flexible active site, lesion-replicating polymerases are much more likely to incorporate the wrong nucleotide into a newly made daughter strand. The mutation rate is typically in the range of 10^{-2} to 10^{-3}. This phenomenon is called **error-prone replication.** By comparison, replicative DNA polymerases are highly intolerant of the geometric distortions imposed on DNA by the incorporation of incorrect nucleotides, and consequently, they incorporate wrong nucleotides with a very low frequency of approximately 10^{-8}. In other words, they copy DNA with a high degree of fidelity.

In *E. coli*, translesion synthesis has been shown to occur under extreme conditions that promote DNA damage, an event termed the **SOS response.** Causes of the SOS response include high doses of UV light and other types of mutagens. Such environmental factors result in the up regulation of several genes that function to repair the DNA lesions, restore replication, and prevent premature cell division. When the SOS response occurs, the cell first attempts to repair the DNA lesions. Any damaged DNA that has not been repaired is replicated by DNA polymerases II, IV, and V, which can replicate over damaged regions. As previously mentioned, this translesion DNA synthesis results in a high rate of mutation. Even though this result may be harmful, it nevertheless allows the bacteria to survive under conditions of extreme environmental stress. Furthermore, the high rate of mutation may provide genetic variability within the bacterial population, so certain cells may become resistant to the harsh conditions.

18.5 REVIEWING THE KEY CONCEPTS

- All species have a variety of DNA repair systems to avoid the harmful effects of mutations (see Table 18.7).
- Photolyase and alkyltransferase can directly repair certain types of damaged bases (see Figure 18.17).
- Base excision repair recognizes and removes a damaged base (see Figure 18.18).
- Nucleotide excision repair removes damaged bases and damaged segments of DNA. Human inherited diseases involve defects in nucleotide excision repair (see Figures 18.19, 18.20).
- Base mismatch repair recognizes a base mismatch and removes a segment of the DNA strand with the incorrect base (see Figure 18.21).
- Double-strand breaks can be repaired by homologous recombination repair and by nonhomologous end joining (see Figures 18.22, 18.23).
- Transcription repair coupling factors can target DNA repair systems to damaged DNA (see Figure 18.24).
- Damaged DNA may be replicated by translesion DNA polymerases that are error-prone.

18.5 COMPREHENSION QUESTIONS

1. The function of photolyase is to repair
   a. double-strand breaks.
   b. apurinic sites.
   c. thymine dimers.
   d. all of the above.
2. Which of the following DNA repair systems may involve the removal of a segment of a DNA strand?
   a. Base excision repair
   b. Nucleotide excision repair
   c. Mismatch repair
   d. All of the above
3. In nucleotide excision repair in *E. coli*, the function of the UvrA/UvrB complex is to
   a. detect DNA damage.
   b. make cuts on both sides of the damage.
   c. remove the damaged piece of DNA.
   d. replace the damaged DNA with undamaged DNA.
4. Double-strand breaks can be repaired by
   a. homologous recombination repair (HRR).
   b. nonhomologous end joining (NHEJ).
   c. nucleotide excision repair (NER).
   d. both a and b.
5. An advantage of translesion DNA polymerases is they can replicate ____________, but a disadvantage is they ____________ .
   a. very quickly, have low fidelity
   b. over damaged DNA, have low fidelity
   c. when resources are limited, are very slow
   d. over damaged DNA, are very slow

KEY TERMS

Page 411. mutation
Page 412. point mutation, base substitution, transition, transversion, silent mutations, missense mutations, nonsense mutations, polarity, frameshift mutations, neutral mutation
Page 414. up promoter mutation, down promoter mutation, wild type, mutant allele, reversion, deleterious mutation, lethal mutation, beneficial mutation, conditional mutants
Page 415. suppressors, suppressor mutations, intragenic suppressor, intergenic suppressor, breakpoint, position effect
Page 417. germ line, germ-line mutation, somatic cells, somatic mutation, genetic mosaic
Page 420. fluctutation test, replica plating, random mutation theory, hot spots
Page 421. spontaneous mutations, induced mutations, depurination, apurinic site
Page 422. deamination, tautomeric shift, tautomers
Page 424. reactive oxygen species (ROS), oxidative stress, oxidative DNA damage, trinucletoide repeat expansion (TNRE), anticipation, dynamic mutation
Page 427. mutagens, nitrous acid, nitrogen mustard, ethyl methanesulfonate (EMS), acridine dyes, proflavin, 5-bromouracil (5BU), 2-aminopurine
Page 428. thymine dimers
Page 429. mutation rate, mutation frequency, Ames test
Page 430. photolyase, photoreactivation, alkyltransferase
Page 431. base excision repair (BER), DNA *N*-glycosylases, AP endonuclease
Page 432. nick translation, nucleotide excision repair (NER)
Page 434. base pair mismatch, mismatch repair system
Page 435. homologous recombination repair (HRR), nonhomologous end joining (NHEJ)
Page 436. transcription-coupled DNA repair, transcription-repair coupling factor (TRCF)
Page 437. translesion synthesis (TLS), error-prone replication, SOS response
Page 439. polar mutation

CHAPTER SUMMARY

- A mutation is a heritable change in the genetic material.

18.1 Effects of Mutations on Gene Structure and Function

- A point mutation is a change in a single base pair. These can be transitions or transversions.
- Silent, missense, nonsense, and frameshift mutations may occur within the coding region of a gene (see Table 18.1, Figure 18.1).
- Mutations may also occur within noncoding regions of a gene and affect gene expression (see Table 18.2).
- Suppressor mutations reverse the phenotypic effects of another mutation. They can be intragenic or intergenic (see Table 18.3).
- Changes in chromosome structure can have a position effect that alters gene expression (see Figures 18.2, 18.3).
- With regard to timing, mutations can occur in the germ line or in somatic cells (see Figures 18.4, 18.5).

18.2 Random Nature of Mutations

- The fluctuation test performed by Luria and Delbrück and the replica plating experiments performed by Lederberg and Lederberg were consistent with the random mutation theory (see Figures 18.6, 18.7).

18.3 Spontaneous Mutations

- Spontaneous mutations result from natural biological or chemical processes, whereas induced mutations are caused by environmental agents (see Table 18.4).
- Three common ways that mutations can arise spontaneously is by depurination, deamination, and tautomeric shifts (see Figures 18.8–18.10).

- Reactive oxygen species (ROS) can cause spontaneous mutations by oxidizing bases in DNA (see Figure 18.11).
- In individuals with trinucleotide repeat expansion (TNRE), a trinucleotide repeat increases above a certain critical size and becomes prone to frequent expansion. This type of mutation, called a dynamic mutation, is responsible for certain types of human diseases. The repeats can expand due to hairpin formation during DNA replication (see Table 18.5, Figure 18.12).

18.4 Induced Mutations

- A mutagen is an agent that can cause a mutation. Researchers have identified many different mutagens that change DNA structure in a variety of ways (see Table 18.6, Figures 18.13–18.15).
- Mutation rate is the likelihood that a new mutation will occur. Mutation frequency is the number of mutant genes divided by the total number of genes in a population.
- The Ames test is used to determine if an agent is a mutagen (see Figure 18.16).

18.5 DNA Repair

- All species have a variety of DNA repair systems to avoid the harmful effects of mutations (see Table 18.7).
- Photolyase and alkyltransferase can directly repair certain types of damaged bases (see Figure 18.17).
- Base excision repair recognizes and removes a damaged base (see Figure 18.18).
- Nucleotide excision repair removes damaged bases and damaged segments of DNA. Human inherited diseases involve defects in nucleotide excision repair (see Figures 18.19, 18.20).
- Base mismatch repair recognizes a base mismatch and removes a segment of the DNA strand with the incorrect base (see Figure 18.21).
- Double-strand breaks can be repaired by homologous recombination repair and by nonhomologous end joining (see Figures 18.22, 18.23).
- Transcription repair coupling factors can target DNA repair systems to damaged DNA (see Figure 18.24).
- Damaged DNA may be replicated by translesion DNA polymerases that are error-prone.

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Mutant tRNAs may act as nonsense and missense suppressors. At the molecular level, explain how you think these suppressors work.

Answer: A suppressor is a second-site mutation that suppresses the phenotypic effects of a first mutation. Intergenic suppressor mutations in tRNA genes can act as nonsense or missense suppressors. For example, let's suppose a first mutation puts a stop codon into a structural gene. A second mutation in a tRNA gene can alter the anticodon region of a tRNA so that the anticodon recognizes a stop codon but inserts an amino acid at this site. A missense suppressor is a mutation in a tRNA gene that changes the anticodon so that it puts in the wrong amino acid at a codon that is not a stop codon. These mutant tRNAs are termed missense tRNAs. For example, a tRNA that normally recognizes glutamic acid may incur a mutation that changes its anticodon sequence so it recognizes a glycine codon instead. Like nonsense suppressors, missense suppressors can be produced by mutations in the anticodon region of tRNAs so the tRNA recognizes an incorrect codon. Alternatively, missense suppressors can also be produced by mutations in aminoacyl-tRNA synthetases that cause them to attach the incorrect amino acid to a tRNA.

S2. If the rate of mutation is 10^{-5} per gene, how many new mutations per gene would you expect in a population of 1 million bacteria?

Answer: If we multiply the mutation rate times the number of bacteria ($10^{-5} \times 10^6$), we obtain a value of 10 new mutations in this gene per million bacteria. This answer is correct, but it is an oversimplification of mutation rate. For any given gene, the mutation rate is based on a probability that an event will occur. Therefore, when we consider a particular population of bacteria, we should be aware that the actual rate of new mutation would vary. Even though the rate may be 10^{-5}, we would not be surprised if a population of 1 million bacteria had 9 or 11 new mutations in this gene instead of the expected number of 10. We would be surprised if it had 5000 new mutations in this gene, because this value would deviate much too far from our expected number.

S3. In the Ames test, several *Salmonella* strains are used that contain different types of mutations within the gene that encodes an enzyme necessary for histidine biosynthesis. These mutations include transversions, transitions, and frameshift mutations. Why do you think it would be informative to test a mutagen with these different strains?

Answer: Different types of mutagens have different effects on DNA structure. For example, if a mutagen caused transversions, an experimenter would want to use a *Salmonella* strain in which a transversion would convert a *his*$^-$ strain into a *his*$^+$ strain. This type of strain would make it possible to detect the effects of the mutagen.

S4. In Chapter 16, we discussed how bacterial genes can be arranged in an operon structure in which a polycistronic RNA contains the coding sequences for two or more genes. For genes in an operon, a relatively short distance occurs between the stop codon in the first gene and the start codon in the next gene. In addition to the ribosome-binding sequence at the beginning of the first gene, a ribosome-binding sequence is also found at the beginning of the second gene and at the beginning of all subsequent genes. After the ribosome has moved past the stop codon in the first gene, it quickly encounters a ribosome-binding site in the next gene. This prevents the complete disassembly of the ribosome as would normally occur after a stop codon and leads to the efficient translation of the second gene.

Jacob and Monod, who studied mutations in the *lac* operon, discovered a mutation in the *lacZ* gene that had an unusual effect. This mutation resulted in a shorter version of the β-galactosidase protein, and it also prevented the expression of the *lacY* gene, even though the *lacY* gene was perfectly normal. Explain this mutation.

Answer: This mutation introduced a stop codon much earlier in the *lacZ* gene sequence. This type of mutation, which affects the translation of a gene downstream in an operon, is referred to as a **polar mutation,** and the phenomenon is called polarity. The explanation is that the early nonsense codon causes the ribosome to disassemble before it has a chance to reach the next gene sequence in the operon. The ribosome-binding sequence that precedes the second gene is efficiently recognized only by a ribosome that has traveled to the normal stop codon in the first gene.

S5. A reverse mutation or a reversion is a mutation that returns a mutant codon back to a codon that gives a wild-type phenotype. At the DNA level, this can be an exact reversion or an equivalent reversion.

GAG (glutamic acid)	First mutation →	GTG (valine)	Exact reversion →	GAG (glutamic acid)
GAG (glutamic acid)	First mutation →	GTG (valine)	Equivalent reversion →	GAA (glutamic acid)
GAG (glutamic acid)	First mutation →	GTG (valine)	Equivalent reversion →	GAT (aspartic acid)

An equivalent reversion produces a protein equivalent to the wild type in structure and function. This can occur in two ways. In some cases, the reversion produces the wild-type amino acid (in this case, glutamic acid), but it uses a different codon than the wild-type gene. Alternatively, an equivalent reversion may substitute an amino acid structurally similar to the wild-type amino acid. In our example, an equivalent reversion has changed valine to an aspartic acid. Because aspartic and glutamic acids are structurally similar—they are acidic amino acids—this type of reversion can restore the wild-type structure and function.

Now here is the question. The template strand within the coding sequence of a gene has the following sequence:

3′–TACCCCTTCGACCCCGGA–5′

This template produces an mRNA,

5′–AUGGGGAAGCUGGGGCCA–3′

that encodes a polypeptide with the following sequence: methionine–glycine–lysine–leucine–glycine–proline.

A mutation changes the template strand to:

3′–TACCCCTACGACCCCGGA–5′

After the first mutation, another mutation occurs to change this sequence again. Would the following second mutations be an exact reversion, an equivalent reversion, or neither?

A. 3′–TACCCCTCCGACCCCGGA–5′

B. 3′–TACCCCTTCGACCCCGGA–5′

C. 3′–TACCCCGACGACCCCGGA–5′

Answer:

A. This is probably an equivalent reversion. The third codon, which encodes a lysine in the normal gene, is now an arginine codon. Arginine and lysine are both basic amino acids, so the polypeptide would probably function normally.

B. This is an exact reversion.

C. The third codon, which is a lysine in the normal gene, has been changed to a leucine codon. It is difficult to say if this would be an equivalent reversion or not. Lysine is a basic amino acid and leucine is a nonpolar, aliphatic amino acid. The protein may still function normally with a leucine at the third codon, or it may function abnormally. You would need to test the function of the protein to determine if this was an equivalent reversion or not.

Conceptual Questions

C1. For each of the following mutations, is it a transition, transversion, addition, or deletion? The original DNA strand is

5′–GGACTAGATAC–3′

(Note: Only the coding DNA strand is shown.)

A. 5′–GAACTAGATAC–3′

B. 5′–GGACTAGAGAC–3′

C. 5′–GGACTAGTAC–3′

D. 5′–GGAGTAGATAC–3′

C2. A gene mutation changes an AT base pair to a GC pair. This causes a gene to encode a truncated protein that is nonfunctional. An organism that carries this mutation cannot survive at high temperatures. Make a list of all the genetic terms that could be used to describe this type of mutation.

C3. What does a suppressor mutation suppress? What is the difference between an intragenic and an intergenic suppressor?

C4. How would each of the following types of mutations affect the amount of functional protein that is expressed from a gene?

A. Nonsense

B. Missense

C. Up promoter mutation

D. Mutation that affects splicing

C5. X-rays strike a chromosome in a living cell and ultimately cause the cell to die. Did the X-rays produce a mutation? Explain why or why not.

C6. Lactose permease is encoded by the *lacY* gene of the *lac* operon. A mutation occurred at codon 64 that changed the normal glycine codon into a valine codon. The mutant lactose permease is unable to function. However, a second mutation, which changes codon 50 from an alanine codon to a threonine codon, is able to restore function. Are the following terms appropriate or inappropriate to describe this second mutation?

A. Reversion

B. Intragenic suppressor

C. Intergenic suppressor

D. Missense mutation

C7. Nonsense suppressors, which are described in Solved Problem S1, tend to be very inefficient at their job of allowing readthrough of a stop codon. How would it affect the cell if they were efficient at their job?

C8. Are each of the following mutations silent, missense, nonsense, or frameshift mutations? The original DNA strand is 5′–ATGGGACTAGATACC–3′. (Note: Only the coding strand is shown; the first codon is methionine.)

A. 5′–ATGGGTCTAGATACC–3′

B. 5′–ATGCGACTAGATACC–3′

C. 5′–ATGGGACTAGTTACC–3′

D. 5′–ATGGGACTAAGATACC–3′

C9. In Chapters 14 through 17, we discussed many sequences that are outside the coding sequence and are important for gene expression. Look up two of these sequences and write them out. Explain how a mutation could change these sequences, thereby altering gene expression.

C10. Explain two ways that a chromosomal rearrangement can cause a position effect.

C11. Is a random mutation more likely to be beneficial or harmful? Explain your answer.

C12. Which of the following mutations could be appropriately described as a position effect?

A. A point mutation at the –10 position in the promoter region prevents transcription.

B. A translocation places the coding sequence for a muscle-specific gene next to an enhancer that is turned on in nerve cells.

C. An inversion flips a gene from the long arm of chromosome 17 (which is euchromatic) to the short arm (which is heterochromatic).

C13. As discussed in Chapter 23, most forms of cancer are caused by environmental agents that produce mutations in somatic cells. Is an individual with cancer considered a genetic mosaic? Explain why or why not.

C14. Discuss the consequences of a germ-line versus a somatic mutation.

C15. Draw and explain how alkylating agents alter the structure of DNA.

C16. Explain how a mutagen can interfere with DNA replication to cause a mutation. Give two examples.

C17. What type of mutation (transition, transversion, or frameshift) would you expect each of the following mutagens to cause?

A. Nitrous acid

B. 5-Bromouracil

C. Proflavin

C18. Explain what happens to the sequence of DNA during trinucleotide repeat expansion (TNRE). If someone was mildly affected with a TNRE disorder, what issues would be important when considering whether to have offspring?

C19. Distinguish between spontaneous and induced mutations. Which are more harmful? Which are avoidable?

C20. Are mutations random events? Explain your answer.

C21. Give an example of a mutagen that can change cytosine to uracil. Which DNA repair system(s) would be able to repair this defect?

C22. If a mutagen causes bases to be removed from nucleotides within DNA, what repair system would fix this damage?

C23. Trinucleotide repeat expansions (TNREs) are associated with several different human inherited diseases. Certain types of TNREs produce a long stretch of glutamines (an amino acid) within the encoded protein. This long stretch of glutamines somehow inhibits the function of the protein, thereby causing a disorder. When a TNRE exerts its detrimental effect by producing a glutamine stretch, are the following statements true or false?

A. The TNRE is within the coding sequence of the gene.

B. The TNRE prevents RNA polymerase from transcribing the gene properly.

C. The trinucleotide sequence is CAG.

D. The trinucleotide sequence is CCG.

C24. With regard to TNRE, what is meant by the term *anticipation*?

C25. What is the difference between the mutation rate and the mutation frequency?

C26. Achondroplasia is a rare form of dwarfism. It is caused by an autosomal dominant mutation within a single gene. Among 1,422,000 live births, the number of babies born with achondroplasia was 31. Among those 31 babies, 18 of them had one parent with achondroplasia. The remaining babies had two unaffected parents. What is the mutation frequency for this disorder among these 1,422,000 babies? What is the mutation rate for achondroplasia?

C27. A segment of DNA has the following sequence:

TTGGATGCTGG
AACCTACGACC

A. What would be the sequence immediately after reaction with nitrous acid? Let the letters H represent hypoxanthine and U represent uracil.

B. Let's suppose this DNA was reacted with nitrous acid. The nitrous acid was then removed, and the DNA was replicated for two generations. What would be the sequences of the DNA products after the DNA had replicated two times? Your answer should contain the sequences of four double helices.

C28. In the treatment of cancer, the basis for many types of chemotherapy and radiation therapy is that mutagens are more effective at killing dividing cells than nondividing cells. Explain why. What are possible harmful side effects of chemotherapy and radiation therapy?

C29. An individual contains a somatic mutation that changes a lysine codon into a glutamic acid codon. Prior to acquiring this mutation, the individual had been exposed to UV light, proflavin, and 5-bromouracil. Which of these three agents would be the most likely to have caused this somatic mutation? Explain your answer.

C30. Which of the following examples is likely to be caused by a somatic mutation?

A. A purple flower has a small patch of white tissue.

B. One child, in a family of seven, is an albino.

C. One apple tree, in a very large orchard, produces its apples 2 weeks earlier than any of the other trees.

D. A 60-year-old smoker develops lung cancer.

C31. How would nucleotide excision repair be affected if one of the following proteins were missing? Describe the condition of the DNA that had been repaired in the absence of the protein.

A. UvrA

B. UvrC

C. UvrD

D. DNA polymerase

C32. During mismatch repair, why is it necessary to distinguish between the template strand and the newly made daughter strand? How is this accomplished?

C33. What are the two main mechanisms by which cells repair double-strand breaks? Briefly describe each one.

C34. With regard to the repair of double-strand breaks, what are the advantages and disadvantages of homologous recombination repair versus nonhomologous end joining?

C35. When DNA *N*-glycosylase recognizes thymine dimers, it detects only the thymine located on the 5′ side of the thymine dimer as being abnormal. Draw and explain the steps whereby a thymine dimer is repaired by the consecutive actions of DNA *N*-glycosylase, AP endonuclease, and DNA polymerase.

C36. Discuss the relationship between transcription and DNA repair. Why is it beneficial to repair actively transcribed DNA more efficiently than nontranscribed DNA?

C37. Three common ways to repair changes in DNA structure are nucleotide excision repair, mismatch repair, and homologous recombination repair. Which of these three mechanisms would be used to fix the following types of DNA changes?

A. A change in the structure of a base caused by a mutagen in a nondividing eukaryotic cell

B. A change in DNA sequence caused by a mistake made by DNA polymerase

C. A thymine dimer in the DNA of an actively dividing bacterial cell

C38. What is the underlying genetic defect that causes xeroderma pigmentosum? How can the symptoms of this disease be explained by the genetic defect?

C39. In *E. coli*, a methyltransferase enzyme encoded by the *dam* gene recognizes the sequence 5′–GATC–3′ and attaches a methyl group to the N^6 position of adenine. *E. coli* strains that have the *dam* gene deleted are known to have a higher spontaneous mutation rate than normal strains. Explain why.

C40. Discuss the similarities and differences between the nucleotide excision repair and mismatch repair systems.

Application and Experimental Questions

E1. The Luria-Delbrück fluctuation test is consistent with the random mutation theory. How would the results have been different if the physiological adaptation hypothesis had been correct?

E2. Explain how the technique of replica plating supports the random mutation theory but conflicts with the physiological adaptation hypothesis. Outline how you would use this technique to show that antibiotic resistance is due to random mutations.

E3. From an experimental point of view, is it better to use haploid or diploid organisms for mutagen testing? Consider the Ames test when preparing your answer.

E4. How would you modify the Ames test to discover physical mutagens? Would it be necessary to add the rat liver extract? Explain why or why not.

E5. During an Ames test, bacteria were exposed to a potential mutagen. Also, as a control, another sample of bacteria was not exposed to the mutagen. In both cases, 10 million bacteria were plated and the following results were obtained:

No mutagen: 17 colonies

With mutagen: 2017 colonies

Calculate the mutation rate in the presence and absence of the mutagen. How much does the mutagen increase the rate of mutation?

E6. Richard Boyce and Paul Howard-Flanders conducted an experiment that provided biochemical evidence that thymine dimers are removed from DNA by a DNA repair system. In their studies, bacterial DNA was radiolabeled so the amount of radioactivity reflected the amount of thymine dimers. The DNA was then subjected to UV light, causing the formation of thymine dimers. When radioactivity was found in the soluble fraction, thymine dimers had been excised from the DNA by a DNA repair system. But when the radioactivity was in the insoluble fraction, the thymine dimers had been retained within the DNA. The following table illustrates some of their results involving a normal strain of *E. coli* and a second strain that was very sensitive to killing by UV light:

Strain	**Treatment**	**Radioactivity in the Insoluble Fraction (cpm)**	**Radioactivity in the Soluble Fraction (cpm)**
Normal	No UV	<100	<40
Normal	UV-treated, incubated 2 hours at 37°C	357	940
Mutant	No UV	<100	<40
Mutant	UV-treated, incubated 2 hours at 37°C	890	<40

Adapted from R. P. Boyce and P. Howard-Flanders (1964) Release of ultraviolet light-induced thymine dimers from DNA in *E. coli* K-12. *Proc Natl Acad Sci USA 51*, 293–300.

Explain the results found in this table. Why is the mutant strain sensitive to UV light?

Questions for Student Discussion/Collaboration

1. In *E. coli*, a variety of mutator strains have been identified in which the spontaneous rate of mutation is much higher than in normal strains. Make a list of the types of abnormalities that could cause a strain of bacteria to become a mutator strain. Which abnormalities do you think would give the highest rate of spontaneous mutation?

2. Discuss the times in a person's life when it would be most important to avoid mutagens. Which parts of a person's body should be most protected from mutagens?

3. A large amount of research is aimed at studying mutation. However, there is not an infinite amount of research dollars. Where would you put your money for mutation research?

A. Testing of potential mutagens

B. Investigating molecular effects of mutagens

C. Investigating DNA repair mechanisms

D. Or some other place

Answers to Comprehension Questions

18.1: b, c, a, d 18.2: b 18.3: c, d, c 18.4: b, b, b, a 18.5: c, d, a, d, b

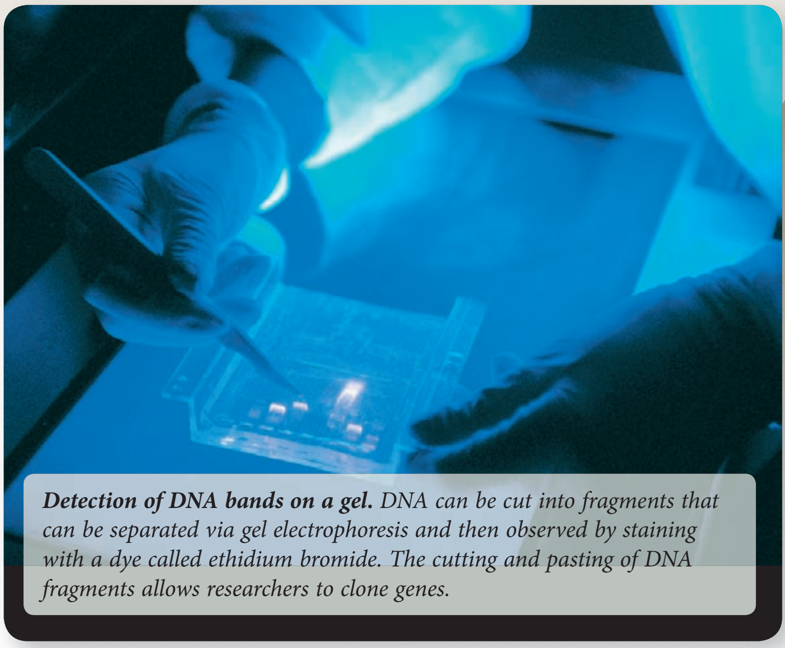
Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

PART V GENETIC TECHNOLOGIES

19


Detection of DNA bands on a gel. *DNA can be cut into fragments that can be separated via gel electrophoresis and then observed by staining with a dye called ethidium bromide. The cutting and pasting of DNA fragments allows researchers to clone genes.*

CHAPTER OUTLINE

19.1 Gene Cloning Using Vectors
19.2 Polymerase Chain Reaction
19.3 DNA Libraries and Blotting Methods
19.4 Methods for Analyzing DNA- and RNA-Binding Proteins
19.5 DNA Sequencing and Site-Directed Mutagenesis

RECOMBINANT DNA TECHNOLOGY

As we learned in Chapter 13, recombination involves the cutting and pasting of DNA segments to produce new arrangements. **Recombinant DNA technology** is the use of in vitro molecular techniques to manipulate fragments of DNA to produce new arrangements. In the early 1970s, the first successes in making recombinant DNA molecules were accomplished independently by two groups at Stanford University: David Jackson, Robert Symons, and Paul Berg; and Peter Lobban and A. Dale Kaiser. Both groups were able to isolate and purify pieces of DNA in a test tube and then covalently link DNA fragments from two different sources. In other words, they constructed molecules called **recombinant DNA molecules.** Shortly thereafter, researchers were able to introduce such recombinant DNA molecules into living cells. Once inside a host cell, the recombinant molecules can be replicated to produce many identical copies of a gene—a process called gene cloning.

Recombinant DNA technology and gene cloning have enabled geneticists to probe relationships between gene sequences and phenotypic consequences and thereby have been fundamental to our understanding of gene structure and function. Most researchers in molecular genetics are familiar with recombinant DNA technology and apply it frequently in their work. Significant practical applications of recombinant DNA technology also have been developed, including exciting advances such as gene therapy, screening for human diseases, recombinant vaccines, and the production of transgenic plants and animals in agriculture, in which a cloned gene from one species is transferred to another species. Transgenic organisms have also been important in basic research.

In this chapter, we will focus primarily on the use of recombinant DNA technology as a way to further our understanding of gene structure and function. We will examine the materials and molecular techniques used in gene cloning, and explore the technique of polymerase chain reaction (PCR), which can make many copies of DNA from a defined region. We will then consider techniques for identifying specific genes or gene products, as well as methods for detecting the binding of proteins to DNA or RNA sequences. Finally, we will examine how scientists analyze and alter DNA sequences through the techniques of DNA sequencing (a method that enables researchers to determine the base sequence of a DNA strand) and site-directed mutagenesis (a procedure that allows researchers to make mutations within a cloned segment of DNA).

In Chapter 20, we will consider many of the practical applications that have arisen as a result of these technologies. Chapters 21 and 22 are devoted to genomics, the molecular analysis of many genes and even the entire genome of a species.

19.1 GENE CLONING USING VECTORS

Molecular biologists want to understand how the molecules within living cells contribute to cell structure and function. Because proteins are the workhorses of cells and because they are the products of genes, many molecular biologists focus their attention on the structure and function of proteins or the genes that encode them. Researchers may focus their efforts on the study of just one or perhaps a few different genes or proteins. At the molecular level, this poses a daunting task. In eukaryotic species, any given cell can express thousands of different proteins, making the study of any single gene or protein akin to a "needle in a haystack" exploration. To overcome this truly formidable obstacle, researchers frequently take the approach of cloning the genes that encode their proteins of interest. The term **gene cloning** refers to the phenomenon of making many copies of a gene. The laboratory methods necessary to clone a gene were devised during the early 1970s. Since then, many technical advances have enabled gene cloning to become a widely used procedure among scientists, including geneticists, cell biologists, biochemists, plant biologists, microbiologists, evolutionary biologists, clinicians, and biotechnologists.

Table 19.1 summarizes some of the common uses of gene cloning. In modern molecular biology, the many uses for gene cloning are remarkable. For this reason, gene cloning has provided the foundation for critical technical advances in a variety of disciplines, including molecular biology, genetics, cell biology, biochemistry, and medicine. In this and the following section, we will examine the two general strategies used to make copies of a gene—the insertion of a gene into a vector that is then propagated in living cells, and cloning via polymerase chain reaction. Later sections in this chapter, as well as Chapters 20 through 22, will consider many of the uses of gene cloning that are described in Table 19.1.

TABLE 19.1

Some Uses of Gene Cloning

Technique	Description
DNA sequencing	Cloned genes provide enough DNA to subject the gene to DNA sequencing (described later in this chapter). The sequence of the gene can reveal the gene's promoter, regulatory sequences, and coding sequence. DNA sequencing is also important in the identification of alleles that cause cancer and inherited human diseases.
Site-directed mutagenesis	A cloned gene can be manipulated to change its DNA sequence. Mutations within genes can help to identify gene sequences such as promoters and regulatory elements. The study of a mutant gene can also help to elucidate its normal function and how its expression may affect the roles of other genes. Mutations in the coding sequence can reveal which amino acids are important for a protein's structure and function.
Gene probes	Labeled DNA strands from a cloned gene can be used as probes for identifying similar or identical genes or RNA. These methods, known as Southern and Northern blot analysis, are described later in this chapter. Probes can also be used to localize genes within intact chromosomes (see Chapter 21). DNA probes are used in techniques such as DNA fingerprinting for identifying crime suspects (see Chapter 25).
Expression of cloned genes	Cloned genes can be introduced into a different cell type or different species. The expression of cloned genes has many uses: **Research** 1. The expression of a cloned gene can help to elucidate its cellular function. 2. The coding sequence of a gene can be placed next to an active promoter and then introduced into a culture of cells that express a large amount of the protein. This greatly aids in the purification of large amounts of protein that may be needed for biochemical or biophysical studies. **Biotechnology** 1. Cloned genes can be introduced into bacteria to make pharmaceutical products such as insulin (see Chapter 20). 2. Cloned genes can be introduced into plants and animals to make transgenic species with desirable traits (see Chapter 20). *Clinical trials* 1. Cloned genes have been used in clinical trials involving gene therapy (see Chapter 20).

Cloning Experiments May Involve Two Kinds of DNA Molecules: Chromosomal DNA and Vector DNA

If a scientist wants to clone a particular gene, a common source of the gene is the chromosomal DNA of the species that carries the gene. For example, if the goal is to clone the rat β-globin gene, this gene is found within the chromosomal DNA of rat cells. To prepare chromosomal DNA, an experimenter first obtains cellular tissue from the organism of interest. The preparation of chromosomal DNA then involves the breaking open of cells and the extraction and purification of the DNA using biochemical techniques such as chromatography and centrifugation (see the Appendix for a description of these techniques).

Let's begin our discussion of gene cloning by considering a recombinant DNA technology in which a gene is removed from its native site within a chromosome and inserted into a smaller segment of DNA known as a **vector**—a small DNA molecule that can replicate independently of host cell chromosomal DNA and produce many identical copies of an inserted gene. The purpose of vector DNA is to act as a carrier of the DNA segment that is to be cloned. (The term *vector* comes from a Latin term meaning carrier.) In cloning experiments, a vector may carry a small segment of chromosomal DNA, perhaps only a single gene. By comparison, a chromosome carries many more genes, perhaps a few hundred or thousand. Like a chromosome, a vector is replicated when it resides within a living cell; a cell that harbors a vector is called a **host cell.** When a vector is replicated within a host cell, the DNA that it carries is also replicated.

The vectors commonly used in gene-cloning experiments were derived originally from two natural sources: plasmids or viruses. Most vectors are **plasmids,** which are small circular pieces of DNA. As discussed in Chapter 9, plasmids are found naturally in many strains of bacteria and occasionally in eukaryotic cells. Many naturally occurring plasmids carry genes that confer resistance to antibiotics or other toxic substances. These plasmids are called resistance plasmids, or **R factors.** Some of the plasmids used in modern cloning experiments were derived from R factors.

Plasmids also contain a DNA sequence, known as an **origin of replication,** that is recognized by the replication enzymes of the host cell and allow it to be replicated. The sequence of the origin of replication determines whether or not the vector can replicate in a particular type of host cell. Some plasmids have origins of replications with a broad host range. Such a plasmid can replicate in the cells of many different species. Alternatively, many vectors used in cloning experiments have a limited host cell range. In cloning experiments, researchers must choose a vector that replicates in the appropriate cell type(s) for their experiments. For example, if researchers want a cloned gene to be propagated in *Escherichia coli,* the vector they employ must have an origin of replication that is recognized by this species of bacterium. The origin of replication also determines the copy number of a plasmid—the number of plasmids typically present in a cell. Some plasmids are said to have strong origins because they can achieve high copy number—perhaps 100 to 200 copies of the plasmid per cell. Others have weaker origins, whereby only one or two plasmids are found per cell.

Commercially available plasmids have been genetically engineered for effective use in cloning experiments. They contain unique sites where geneticists can easily insert pieces of DNA. Another useful feature of cloning vectors is they often contain resistance genes that provide host cells with the ability to grow in the presence of a toxic substance. Such a gene is called a **selectable marker,** because the expression of the gene selects for the growth of the host cells. Many selectable markers are genes that confer antibiotic resistance to the host cell. For example, the gene *amp*^{R} encodes an enzyme known as β-lactamase. This enzyme degrades ampicillin, an antibiotic that normally kills bacteria. Bacteria containing the *amp*^{R} gene can grow on media containing ampicillin because they can degrade it. In a cloning experiment where the *amp*^{R} gene is found within the plasmid, the growth of cells in the presence of ampicillin identifies bacteria that contain the plasmid. These bacteria can grow and form visible colonies on solid growth media. In contrast, those cells that do not contain the plasmid are ampicillin-sensitive and do not grow.

An alternative type of vector used in cloning experiments is a viral vector. As discussed in Chapter 10, viruses can infect living cells and propagate themselves by taking control of the host cell's metabolic machinery. When a chromosomal gene is inserted into a viral genome, the gene is replicated whenever the viral DNA is replicated. Therefore, viruses can be used as vectors to carry other pieces of DNA. When a virus is used as a vector, the researcher may analyze viral plaques rather than bacterial colonies. The characteristics of viral plaques are described in Chapter 10 (see Figure 10.7).

Molecular biologists may choose from hundreds of different vectors to use in their cloning experiments. Table 19.2 provides a general description of several different types of vectors that are commonly used to clone small segments of DNA. In addition, other types of vectors, such as cosmids, bacterial artificial chromosomes (BACs), and yeast artificial chromosomes (YACs), are used to clone large pieces of DNA. These vectors are described in detail in Chapter 21. Vectors designed to introduce genes into plants and animals are discussed in Chapter 20.

Enzymes Are Used to Cut DNA into Pieces and Join the Pieces Together

A key step in a cloning experiment is the insertion of chromosomal DNA into a plasmid or viral vector. This requires the

TABLE 19.2

Some Vectors Used in Cloning Experiments

Example	Type	Description
pBluescript	Plasmid	A type of vector like the one shown in Figure 19.2. It is used to clone small segments of DNA and propagate them in *E. coli.*
YEp24	Plasmid	This plasmid is an example of a **shuttle vector,** which can replicate in two different host species, *E. coli* and *Saccharomyces cerevisiae.* It carries origins of replication for both species.
λgt11	Viral	This vector is derived from the bacteriophage λ, which is described in Chapter 10. λgt11 also contains a promoter from the *lac* operon. When fragments of DNA are cloned next to this promoter, the DNA is expressed in *E. coli.* This is an example of an **expression vector.** An expression vector is designed to clone the coding sequence of genes so they are transcribed and translated correctly.
SV40	Viral	This virus naturally infects mammalian cells. Genetically altered derivatives of the SV40 viral DNA are used as vectors for the cloning and expression of genes in mammalian cells that are grown in the laboratory.
Baculovirus	Viral	This virus naturally infects insect cells. In a laboratory, insect cells can be grown in liquid media. Unlike many other types of eukaryotic cells, insect cells often express large amounts of proteins that are encoded by cloned genes. When researchers want to make a large amount of a protein, they can clone the gene that encodes the protein into baculovirus and then purify the protein from insect cells.

cutting and pasting of DNA fragments. To cut DNA, researchers use enzymes known as **restriction endonucleases,** or **restriction enzymes.** The restriction enzymes used in cloning experiments bind to a specific base sequence and then cleave the DNA backbone at two defined locations, one in each strand. Proposed by Werner Arber in the 1960s and discovered by Hamilton Smith and Daniel Nathans in the 1970s, restriction enzymes are made naturally by many different species of bacteria and protect bacterial cells from invasion by foreign DNA, particularly that of bacteriophages. Restriction enzymes are used by bacteria to cut and inactivate the DNA of invading bacteriophages. A bacterium protects its own DNA from digestion by restriction enzymes by modifying the recognition site, usually by methylation.

Researchers have isolated and purified restriction enzymes from many bacterial species and now use them in their cloning experiments. **Figure 19.1** shows an example of using a restriction enzyme, called *Eco*RI, in producing a recombinant DNA molecule. Certain types of restriction enzymes are useful in cloning because they digest DNA into fragments with “sticky ends.” As shown in Figure 19.1, the sticky ends are single-stranded regions of DNA that can hydrogen bond to a complementary sequence of DNA from a different source. The ends of two different DNA

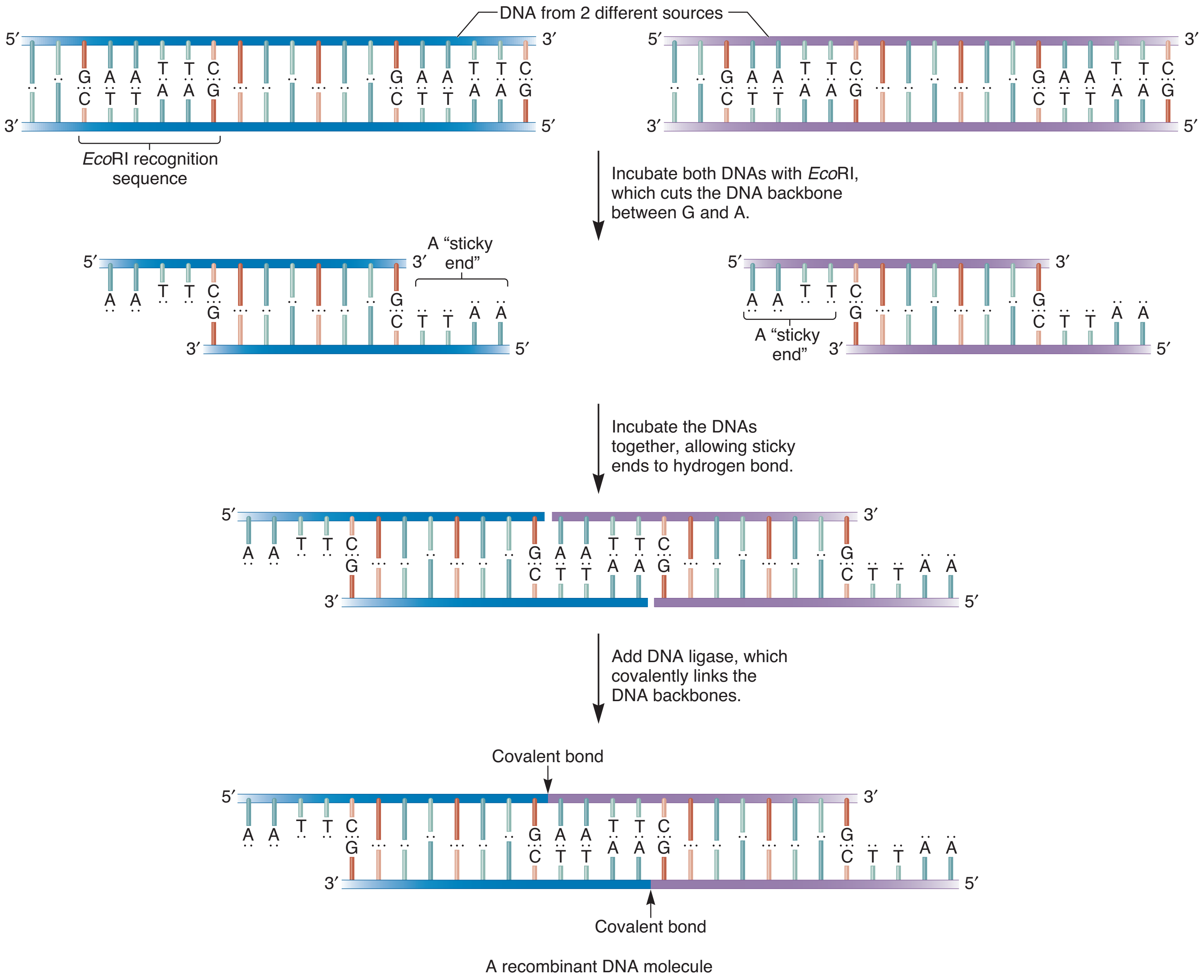

DNA from 2 different sources
5′
3′
G A A T T C
C T T A A G
*Eco*RI recognition sequence
Incubate both DNAs with *Eco*RI, which cuts the DNA backbone between G and A.
A “sticky end”
Incubate the DNAs together, allowing sticky ends to hydrogen bond.
Add DNA ligase, which covalently links the DNA backbones.
Covalent bond
Covalent bond
A recombinant DNA molecule

FIGURE 19.1 **The action of a restriction enzyme and the production of recombinant DNA.** The restriction enzyme *Eco*RI binds to a specific sequence, in this case 5′–GAATTC–3′. It then cleaves the DNA backbone between G and A, producing DNA fragments. The single-stranded ends of different DNA fragments can hydrogen bond with each other, because they have complementary sequences. The enzyme DNA ligase then catalyzes the formation of covalent bonds in the DNA backbones of the fragments.

Concept check: *Prior to the action of DNA ligase, how many hydrogen bonds are holding these two DNA fragments together?*

pieces hydrogen bond to each other because of their complementary sticky ends.

The hydrogen bonding between the sticky ends of DNA fragments promotes a temporary interaction between the two fragments. However, this interaction is not stable because it involves only a few hydrogen bonds between complementary bases. How can this interaction be made more permanent? The answer is that the sugar-phosphate backbones within the DNA strands must be covalently linked together. This linkage is catalyzed by **DNA ligase.** Figure 19.1 illustrates the action of DNA ligase, which catalyzes covalent bond formation in the sugar-phosphate backbones of both DNA strands after the sticky ends have hydrogen bonded with each other.

Currently, several hundred different restriction enzymes from various bacterial species have been identified and are available commercially to molecular biologists. **Table 19.3** gives a few examples. Restriction enzymes usually recognize sequences that are **palindromic;** that is, the sequence in one strand is identical when read in the opposite direction in the complementary strand. For example, the sequence recognized by *Eco*RI is 5′–GAATTC–3′ in the top strand. Read in the opposite direction in the bottom strand, this sequence is also 5′–GAATTC–3′.

TABLE **19.3**

Some Restriction Enzymes Used in Gene Cloning

Restriction Enzyme*	Bacterial Source	Sequence Recognized†
*Bam*HI	*Bacillus amyloliquefaciens* H	↓ 5′–GGATCC–3′ 3′–CCTAGG–5′ ↑
*Cla*I	*Caryophanon latum*	↓ 5′–ATCGAT–3′ 3′–TAGCTA–5′ ↑
*Eco*RI	*E. coli* RY13	↓ 5′–GAATTC–3′ 3′–CTTAAG–5′ ↑
*Nae*I	*Nocardia aerocolonigenes*	↓ 5′–GCCGGC–3′ 3′–CGGCCG–5′ ↑
*Pst*I	*Providencia stuartii*	↓ 5′–CTGCAG–3′ 3′–GACGTC–5′ ↑
*Sac*I	*Streptomyces achromogenes*	↓ 5′–GAGCTC–3′ 3′–CTCGAG–5′ ↑

*Restriction enzymes are named according to the species in which they are found. The first three letters are italicized because they indicate the genus and species names. Because a species may produce more than one restriction enzyme, the enzymes are designated I, II, III, and so on, to indicate the order in which they were discovered in a given species. Some restriction enzymes, like *Eco*RI, produce a sticky end with a 5′ overhang (see Figure 19.1), whereas others, such as *Pst*I, produce a 3′ overhang. However, not all restriction enzymes cut DNA to produce sticky ends. For example, the enzyme *Nae*I cuts DNA to produce blunt ends.

†The arrows show the locations in the upper and lower DNA strands where the restriction enzymes cleave the DNA backbone. A complete list of restriction enzymes can be found at http://rebase.neb.com/rebase/rebase.html.

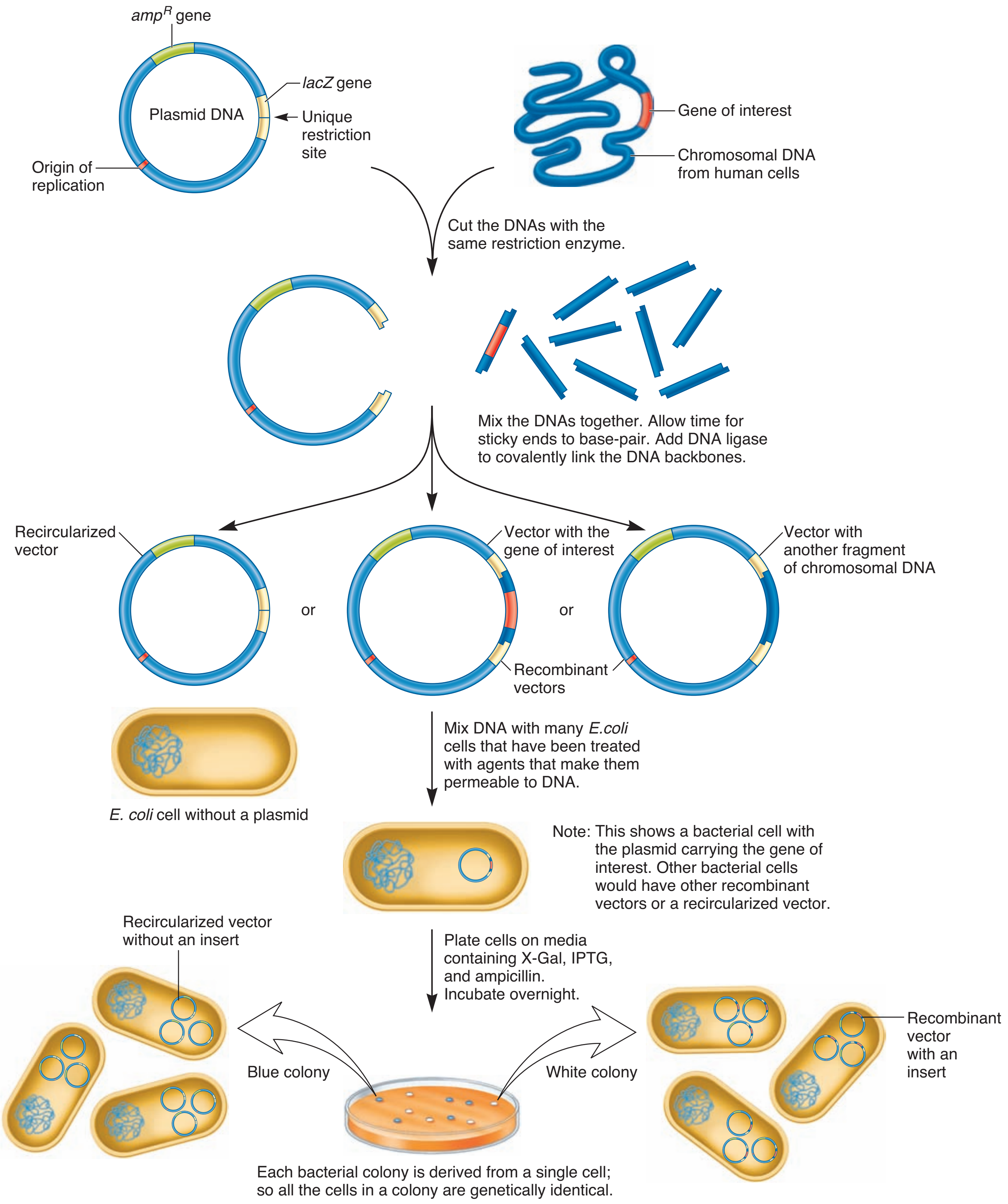
Gene Cloning Involves the Insertion of DNA Fragments into Vectors, Which Are Then Propagated Within Host Cells

Now that we are familiar with the materials, let's outline the general strategy followed in a typical cloning experiment. The goal of the procedure shown in **Figure 19.2** is to clone a chromosomal gene of interest into a plasmid vector that already carries the *amp*^{R} gene. To begin this experiment, the chromosomal DNA is isolated and digested with a restriction enzyme. This enzyme cuts the chromosomes into many small fragments. The plasmid DNA is also cut with the same restriction enzyme. However, the plasmid has only one unique site for the restriction enzyme. After cutting, the plasmid has two ends that are complementary to the sticky ends of the chromosomal DNA fragments. The digested chromosomal DNA and plasmid DNA are mixed together and incubated under conditions that promote the binding of these complementary sticky ends.

DNA ligase is then added to catalyze the covalent linkage between DNA fragments. In some cases, the two ends of the vector simply ligate back together, restoring the vector to its original structure. This is called a recircularized vector. In other cases, a fragment of chromosomal DNA may become ligated to both ends of the vector. In this way, a segment of chromosomal DNA has been inserted into the vector. The vector containing a piece of chromosomal DNA is referred to as a **recombinant vector.** A recombinant vector may carry the gene of interest or it may contain another fragment of chromosomal DNA.

Following ligation, the DNA is introduced into living cells treated with agents that render them permeable to DNA molecules. Cells that can take up DNA from the extracellular medium are called **competent cells.** This step in the procedure is commonly called **transformation** (see Chapter 9). In the experiment shown in Figure 19.2, a plasmid is introduced into bacterial cells that were originally sensitive to ampicillin. The bacteria are then streaked onto plates containing bacterial growth medium and ampicillin. A bacterium that has taken up a plasmid carrying the *amp*^{R} gene continues to divide and forms a bacterial colony containing tens of millions of cells. Because each cell within a single colony is derived from the same original cell, all cells within a colony contain the same type of plasmid DNA.

In the experiment shown in Figure 19.2, how can the experimenter distinguish between bacterial colonies that contain a recircularized vector versus those with a recombinant vector? As shown here, the chromosomal DNA has been inserted into a region of the vector that contains the *lacZ* gene, which encodes the enzyme β-galactosidase. The insertion of chromosomal DNA into the vector disrupts the *lacZ* gene so it cannot produce a functional enzyme. By comparison, a recircularized vector has a functional


ampR gene
lacZ gene
Plasmid DNA
Unique restriction site
Origin of replication
Gene of interest
Chromosomal DNA from human cells
Cut the DNAs with the same restriction enzyme.
Mix the DNAs together. Allow time for sticky ends to base-pair. Add DNA ligase to covalently link the DNA backbones.
Recircularized vector
or
Vector with the gene of interest
or
Vector with another fragment of chromosomal DNA
Recombinant vectors
E. coli cell without a plasmid
Mix DNA with many E.coli cells that have been treated with agents that make them permeable to DNA.
Note: This shows a bacterial cell with the plasmid carrying the gene of interest. Other bacterial cells would have other recombinant vectors or a recircularized vector.
Recircularized vector without an insert
Plate cells on media containing X-Gal, IPTG, and ampicillin. Incubate overnight.
Blue colony
White colony
Recombinant vector with an insert
Each bacterial colony is derived from a single cell; so all the cells in a colony are genetically identical.

ONLINE ANIMATION

INTERACTIVE EXERCISE

FIGURE 19.2 **The steps in gene cloning.** Note: X-Gal refers to the colorless compound 5-bromo-4-chloro-3-indolyl-β-D-galactoside. IPTG is an acronym for isopropyl-β-D-thiogalactopyranoside, which is a nonmetabolizable lactose analog that can induce the *lac* promoter.

Concept check: *Explain the role of the selectable marker gene in this experiment.*

lacZ gene. The functionality of *lacZ* is determined by providing the growth medium with a colorless compound, X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), which is cleaved by β-galactosidase into a blue dye. Bacteria grown in the presence of X-Gal and IPTG (an inducer of the *lacZ* gene) form blue colonies if they have a functional *lacZ* gene and white colonies if they do not. In this experiment, therefore, bacterial colonies containing recircularized vectors form blue colonies, whereas colonies containing recombinant vectors are white.

In the example of Figure 19.2, one of the white colonies contains cells with a recombinant vector that carries a human gene of interest; the segment containing the human gene is shown in red. The goal of gene cloning is to produce an enormous number of copies of a recombinant vector that carries the gene of interest. During transformation, a single bacterial cell usually takes up a single copy of a recombinant vector. However, two subsequent events lead to the amplification of the cloned gene. First, because the vector has an origin of replication, the bacterial host cell replicates the recombinant vector to produce many identical copies per cell. Second, the bacterial cells divide approximately every 20 minutes. Following overnight growth, a population of many millions of bacteria is obtained. Each of these bacterial cells contains many copies of the cloned gene. For example, a bacterial colony may be composed of 10 million cells, with each cell containing 50 copies of the recombinant vector. Therefore, this bacterial colony would contain 500 million copies of the cloned gene!

The preceding description has acquainted you with the steps required to clone a gene. A misleading aspect of Figure 19.2 is that the digestion of the chromosomal DNA with restriction enzymes appears to yield only a few DNA fragments, one of which contains the gene of interest. In an actual cloning experiment, however, the digestion of the chromosomal DNA with a restriction enzyme produces tens of thousands of different pieces of chromosomal DNA, not just a few. Later, we will consider methods of identifying bacterial colonies containing the specific gene that a researcher wants to clone.

Recombinant DNA technology can also be used to clone fragments of DNA that do not code for genes. For example, sequences such as telomeres, centromeres, and highly repetitive sequences have been cloned by this procedure.

cDNA Can Be Made from mRNA via Reverse Transcriptase

In the discussion of gene cloning described earlier in Figure 19.2, chromosomal DNA and plasmid DNA were the materials used to clone genes. Alternatively, a sample of RNA can provide a starting point for cloning DNA. As described in Chapter 10, the enzyme **reverse transcriptase** can use RNA as a template to make a complementary strand of DNA. This enzyme is encoded in the genome of retroviruses and provides a way for retroviruses to copy their RNA genome into DNA molecules that integrate into the host cell's chromosomes.

Researchers can use purified reverse transcriptase in a strategy for cloning genes, using mRNA as the starting material (**Figure 19.3**). To begin this experiment, mRNA is purified from a sample of cells. The mRNA is mixed with primers composed of a string of thymine-containing nucleotides. This short strand of DNA, or **oligonucleotide,** is called a poly-dT primer. Because eukaryotic mRNAs contain a polyA tail, poly-dT primers are complementary to the 3′ end of mRNAs. Reverse transcriptase and deoxyribonucleotides (dNTPs) are then added to

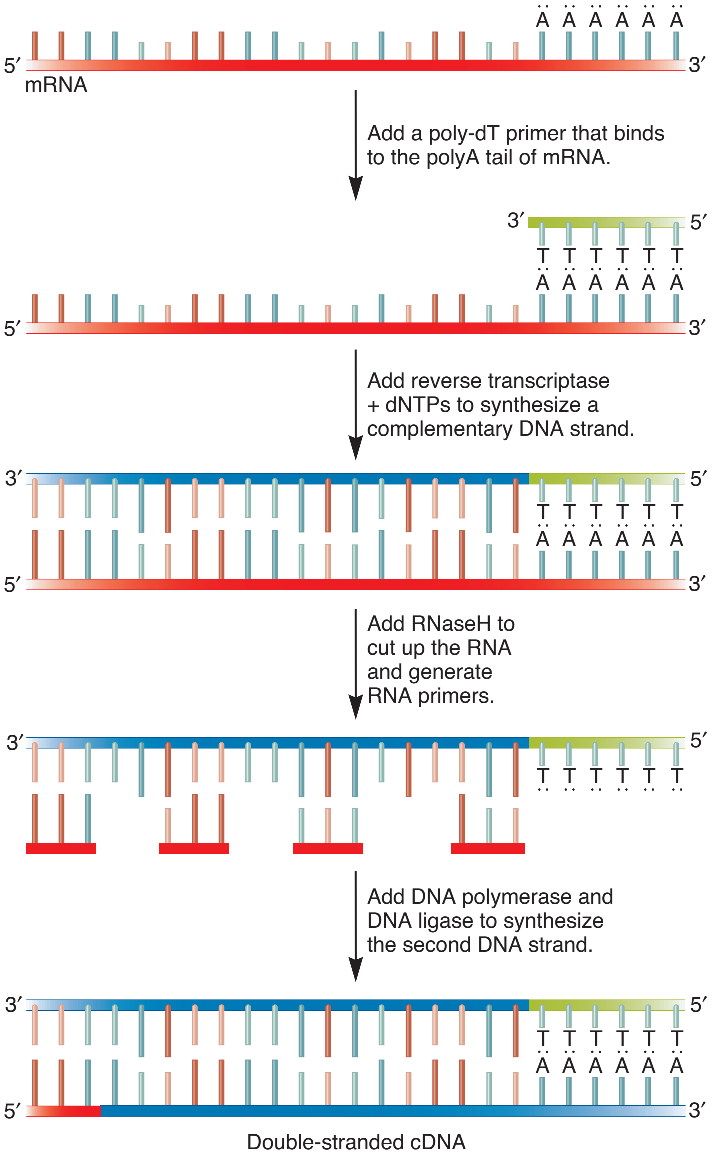

A A A A A A
5′
3′
mRNA
Add a poly-dT primer that binds to the polyA tail of mRNA.
3′
5′
T T T T T T
A A A A A A
5′
3′
Add reverse transcriptase + dNTPs to synthesize a complementary DNA strand.
3′
5′
T T T T T T
A A A A A A
5′
3′
Add RNaseH to cut up the RNA and generate RNA primers.
3′
5′
T T T T T T
Add DNA polymerase and DNA ligase to synthesize the second DNA strand.
3′
5′
T T T T T T
A A A A A A
5′
3′
Double-stranded cDNA

FIGURE 19.3 **Synthesis of cDNA.** A poly-dT primer anneals to the 3′ end of eukaryotic mRNAs. Reverse transcriptase then catalyzes the synthesis of a complementary DNA strand (cDNA). RNaseH digests the mRNA into short pieces that are used as primers by DNA polymerase I to synthesize the second DNA strand. The 5′ to 3′ exonuclease function of DNA polymerase I removes all of the RNA primers except the one at the 5′ end (because there is no primer upstream from this site). This RNA primer can be removed by the subsequent addition of an RNase. After the double-stranded cDNA is made, it can then be inserted into a vector as described later in Figure 19.9b.

Concept check: Explain the meaning of the name "reverse transcriptase."

make a DNA strand that is complementary to the mRNA. Next, RNaseH, DNA polymerase, and DNA ligase are added. RNaseH is an enzyme that partially digests the RNA, generating short RNAs that are used as primers by DNA polymerase to make a second DNA strand that is complementary to the strand made by reverse transcriptase. Finally, DNA ligase seals any nicks in this second DNA strand. When DNA is made from RNA as the starting material, the DNA is called **complementary DNA (cDNA).** The term originally referred to the single strand of DNA that is complementary to the RNA template. However, cDNA now refers to any DNA, whether it is single- or double-stranded, that is made using RNA as the starting material.

Why is cDNA cloning useful? From a research perspective, an important advantage of cDNA is that it lacks introns, which are often found in eukaryotic genes. Because introns can be quite large, it is much simpler to insert cDNAs into vectors if researchers want to focus their attention on the coding sequence of a gene. For example, if the primary goal is to determine the coding sequence of a structural gene, a researcher inserts cDNA into a vector and then determines the DNA sequence of the insert, as described later in this chapter. Similarly, if a scientist wants to express an encoded protein of interest in a cell that does not splice out the introns properly (e.g., in a bacterial cell), it is necessary to make cDNA clones of the respective gene.

19.1 REVIEWING THE KEY CONCEPTS

- Gene cloning has many uses, including DNA sequencing, site-directed mutagenesis, the use of genes as probes, and the expression of cloned genes (see Table 19.1).
- Cloning vectors are derived from plasmids or viruses (see Table 19.2).
- Restriction enzymes, also called restriction endonucleases, cut chromosomal DNA and vector DNA to produce sticky ends that can hydrogen bond with each other. DNA ligase is needed to make a covalent link in the DNA backbone (see Figure 19.1, Table 19.3).
- Gene cloning using vectors involves the insertion of the gene into a vector and then its propagation in a living cell such as *E. coli* (see Figure 19.2).
- Complementary DNA (cDNA) can be made via reverse transcriptase by starting with mRNA (see Figure 19.3).

19.1 COMPREHENSION QUESTIONS

1. Which of the following may be used as a vector in a gene cloning experiment?
   a. mRNA
   b. Plasmid
   c. Virus
   d. Both b and c
2. The restriction enzymes used in gene-cloning experiments ___________, which generates sticky ends that can ___________.
   a. cut the DNA, enter bacterial cells
   b. cut the DNA, hydrogen bond with complementary sticky ends
   c. methylates DNA, enter bacterial cells
   d. methylates DNA, hydrogen bond with complementary sticky ends
3. Which of the following is the proper order of steps in a gene-cloning experiment involving vectors?
   1. Add DNA ligase.
   2. Incubate the chromosomal DNA and the vector DNA with a restriction enzyme.
   3. Introduce the DNA into living cells.
   4. Mix the chromosomal DNA and vector DNA together.
   a. 1, 2, 3, 4
   b. 2, 3, 1, 4
   c. 2, 4, 1, 3
   d. 1, 2, 4, 3
4. The function of reverse transcriptase is to
   a. copy RNA into DNA.
   b. copy DNA into RNA.
   c. translate RNA into protein.
   d. translate DNA into protein.

19.2 POLYMERASE CHAIN REACTION

In our previous discussions of gene cloning, the DNA of interest was inserted into a vector, which then was introduced into a host cell. The replication of the vector within the host cell, and the proliferation of the host cells, led to the production of many copies of the DNA. Another way to copy DNA, without the aid of vectors and host cells, is a technique called **polymerase chain reaction (PCR),** which was developed by Kary Mullis in 1985. In this section, we begin with a general description of PCR and then examine how it can be used to quantitate the amount of DNA or RNA in a biological sample.

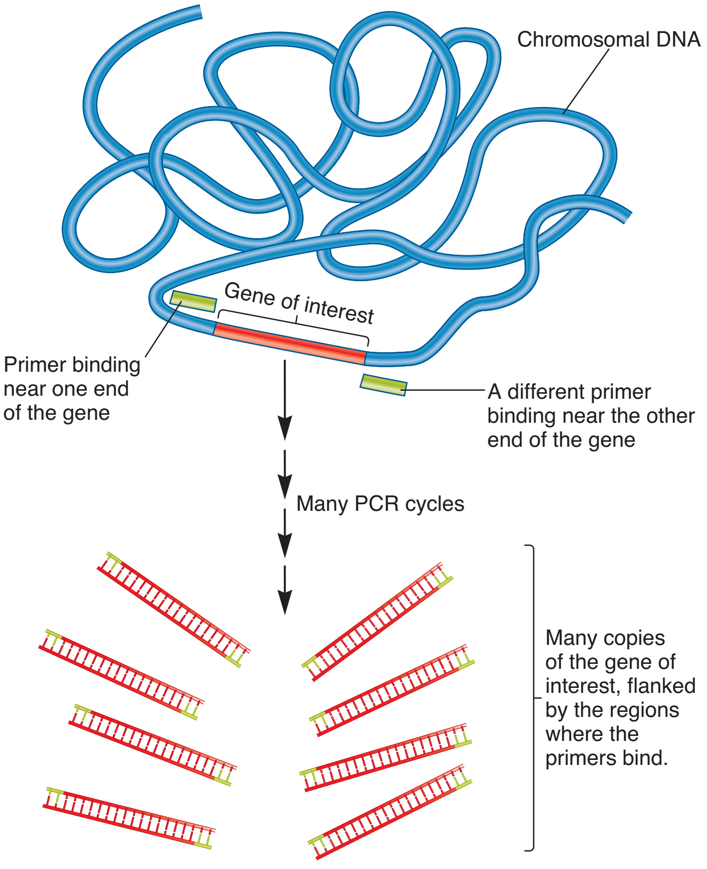
Each Cycle of PCR Involves Three Steps: Denaturation, Primer Annealing, and Primer Extension

The PCR method is used to make large of amounts of DNA in a defined region that is flanked by two **primers.** The primers are oligonucleotides that are about 15 to 20 nucleotides in length. As shown in **Figure 19.4a**, the starting material of a PCR experiment can be a complex mixture of DNA. The two primers bind to specific sites in the DNA because their bases are complementary to these sites. The end result of PCR is that the region that is flanked by the primers, which contains the gene of interest, is amplified. The term *amplification* means that many copies of the region have been made.

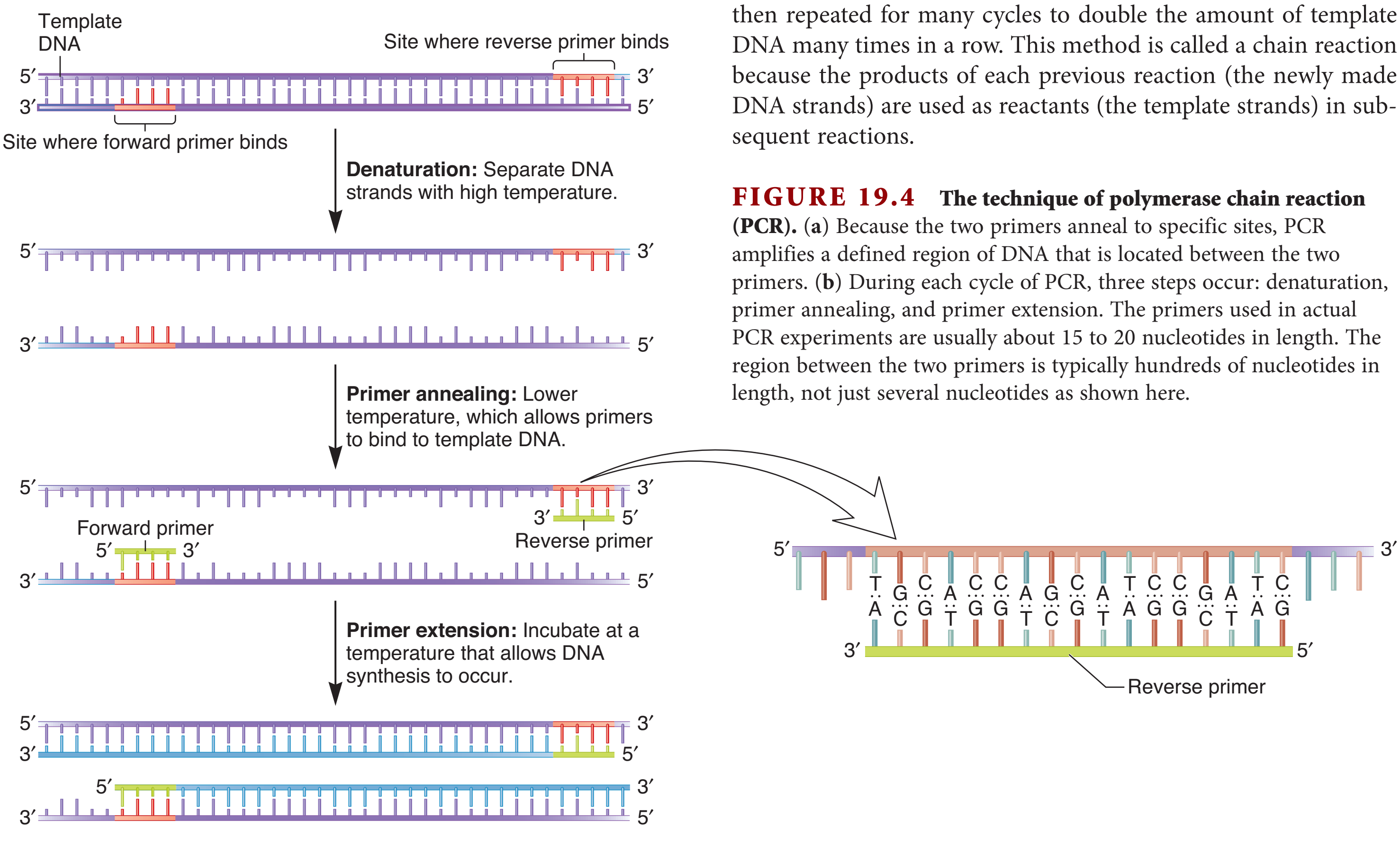
The primers are key reagents in a PCR experiment. They are made chemically. The primers are not typically made in research or clinical laboratories, but instead are made at university or industrial facilities. Researchers or clinicians simply order the primers they need with a specified sequence. How does someone choose the primer sequences? In most cases, PCR is conducted on DNA samples in which a scientist already knows the DNA sequences that flank the region of interest. Because the entire genome of many species has already been determined, sequence information can come from the genome database of the species of interest. For example, if you wanted to clone a human gene or a portion of human gene, such as the β-globin gene, you would look up that gene sequence in the human genome database and decide where you want your primers to bind. You would then order primers with sequences that are complementary to those sites.

Let's now consider the reagents needed for a PCR experiment. The two reagents we have already considered in Figure 19.4a are the primers and a sample of DNA that contains a gene of interest, which is called **template DNA.** In addition, deoxyribonucleoside triphosphates (dNTPs) are needed, as is a thermally stable form of a DNA polymerase such as ***Taq* polymerase,** which is isolated from the bacterium *Thermus aquaticus.* A thermostable form of DNA polymerase is necessary because PCR involves heating steps that inactivate most other natural forms of DNA polymerase (which are thermolabile, or readily denatured by heat).

As outlined in **Figure 19.4b**, PCR involves three steps: denaturation, primer annealing, and primer extension. To make copies of the DNA, the template DNA is first denatured by heat treatment, causing the strands to separate (also see Online Animation). As the temperature is lowered, oligonucleotide primers bind to the DNA in a process called **annealing.** After the primers have annealed, *Taq* polymerase catalyzes the synthesis of complementary DNA strands in the 5′ to 3′ direction, starting at the primers. This process, called **primer extension,** doubles the amount of the template DNA. The three steps shown in Figure 19.4b constitute one cycle of a PCR reaction. The sequential process of denaturation—primer annealing—primer extension is then repeated for many cycles to double the amount of template DNA many times in a row. This method is called a chain reaction because the products of each previous reaction (the newly made DNA strands) are used as reactants (the template strands) in subsequent reactions.

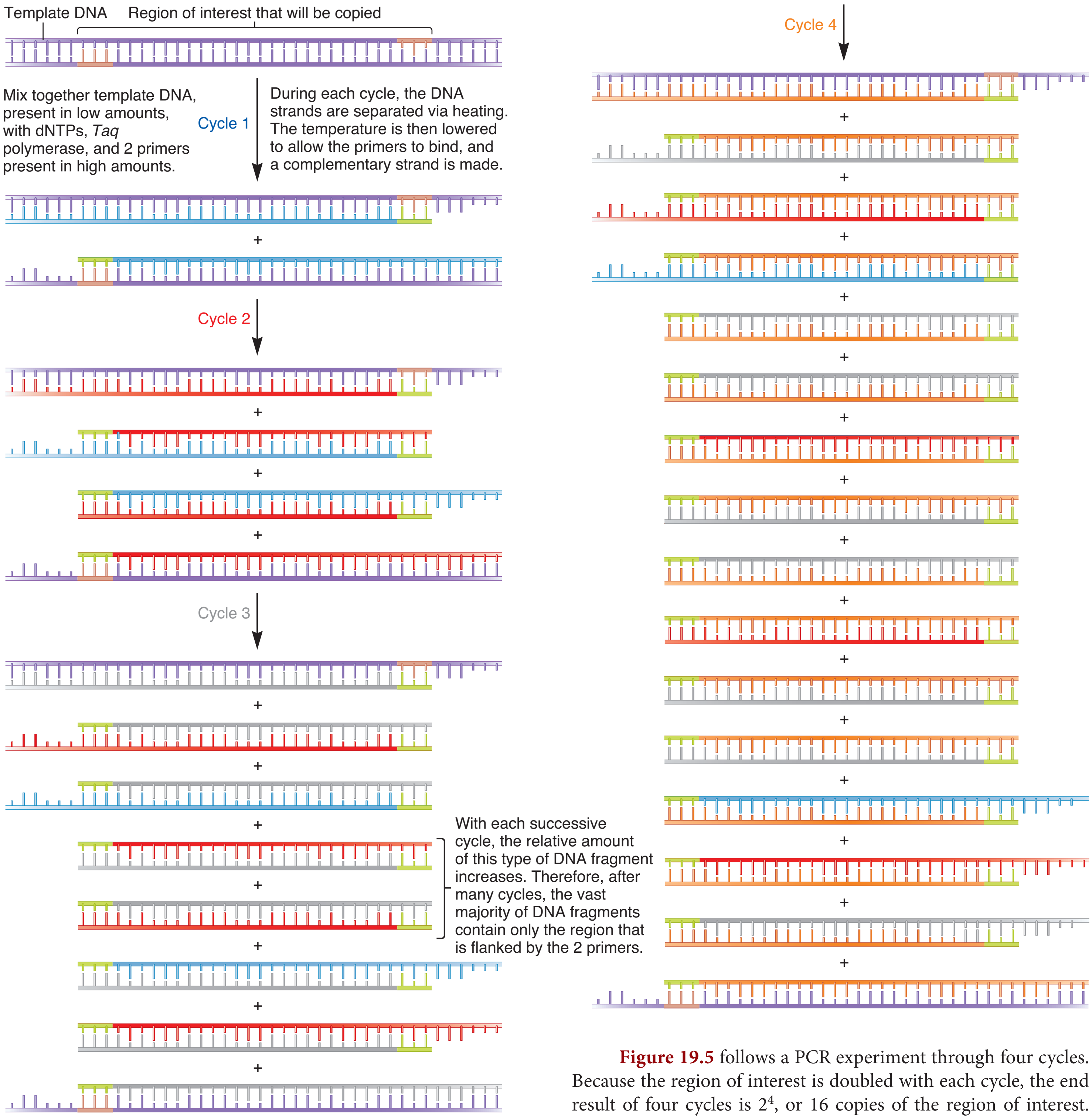

Chromosomal DNA
Gene of interest
Primer binding near one end of the gene
A different primer binding near the other end of the gene
Many PCR cycles
Many copies of the gene of interest, flanked by the regions where the primers bind.

(a) The outcome of a PCR experiment


Template DNA
Site where reverse primer binds
5′
3′
3′
5′
Site where forward primer binds
Denaturation: Separate DNA strands with high temperature.
Primer annealing: Lower temperature, which allows primers to bind to template DNA.
Forward primer
Reverse primer
Primer extension: Incubate at a temperature that allows DNA synthesis to occur.
T G C A C C A G C A T C C G A T C
A C G T G G T C G T A G G C T A G
Reverse primer

(b) The 3 steps of a PCR cycle

FIGURE 19.4 The technique of polymerase chain reaction (PCR). (**a**) Because the two primers anneal to specific sites, PCR amplifies a defined region of DNA that is located between the two primers. (**b**) During each cycle of PCR, three steps occur: denaturation, primer annealing, and primer extension. The primers used in actual PCR experiments are usually about 15 to 20 nucleotides in length. The region between the two primers is typically hundreds of nucleotides in length, not just several nucleotides as shown here.


Template DNA
Region of interest that will be copied
Mix together template DNA, present in low amounts, with dNTPs, *Taq* polymerase, and 2 primers present in high amounts.
Cycle 1
During each cycle, the DNA strands are separated via heating. The temperature is then lowered to allow the primers to bind, and a complementary strand is made.
+
Cycle 2
Cycle 3
Cycle 4
With each successive cycle, the relative amount of this type of DNA fragment increases. Therefore, after many cycles, the vast majority of DNA fragments contain only the region that is flanked by the 2 primers.

ONLINE ANIMATION

FIGURE 19.5 The technique of PCR carried out for four cycles. During each cycle, oligonucleotides that are complementary to the ends of the targeted DNA sequence (green) bind to the DNA and act as primers for the synthesis of this DNA region. Note: The DNA template strands are purple, and those strands made during the first, second, third, and fourth cycles are blue, red, gray, and orange, respectively.

Concept check: After four cycles of PCR, which type of PCR product predominates? Explain why.

Figure 19.5 follows a PCR experiment through four cycles. Because the region of interest is doubled with each cycle, the end result of four cycles is 2^4, or 16 copies of the region of interest. Because the starting material contains long strands of chromosomal DNA, some of the products of a PCR experiment contain the region of interest plus some additional DNA at either end. However, after many cycles, the products that contain only the region of interest greatly predominate the mixture.

PCR is carried out in a thermal reactor, known as a **thermocycler,** that automates the timing of each cycle. The experimenter mixes the DNA sample, dNTPs, *Taq* polymerase, and an excess amount of primers together in a single tube. The tube is placed in a thermocycler, and the experimenter sets the machine to operate within a defined temperature range and number of cycles. During each cycle, the thermocycler increases the temperature to denature the DNA strands and then lowers the temperature to allow primer annealing and extension to take place. Typically, each cycle lasts 2 to 3 minutes and is then repeated. A typical PCR run is likely to involve 20 to 30 cycles of replication and takes a couple of hours to complete. The PCR technique can amplify the amount of DNA by a staggering amount. After 20 cycles of amplification, the intervening region between the two primers increases 2^{20}-fold, which is approximately a million-fold (assuming 100% efficiency)!

An important advantage of PCR is that it can amplify a particular region of DNA from a very complex mixture of template DNA. For example, if a researcher uses two primers that anneal to the human β-globin gene, PCR can amplify just the β-globin gene from a DNA sample that contains all of the human chromosomes!

Alternatively, PCR can be used to amplify a sample of chromosomal DNA nonspecifically. As discussed in Chapter 25, this approach is used in DNA fingerprinting analysis. A nonspecific approach uses a mixture of short PCR primers with many different random sequences. These primers anneal randomly throughout the genome and amplify most of the chromosomal DNA. Nonspecific DNA amplification is used to increase the total amount of DNA in very small samples, such as blood stains found at crime scenes.

Reverse Transcriptase PCR Can Be Used to Amplify RNA

PCR is also used to detect and quantitate the amount of specific RNAs in living cells. To accomplish this goal, RNA is isolated from a sample and mixed with deoxynucleotides, reverse transcriptase, and a primer that binds near the 3′ end of the RNA of interest (**Figure 19.6**). This generates a single-stranded cDNA, which then can be used as template DNA in a conventional PCR reaction. The end result is that the RNA has been amplified to produce many copies of double-stranded cDNA.

This method, called **reverse transcriptase PCR,** is extraordinarily sensitive. Reverse transcriptase PCR can detect the expression of small amounts of RNA from a single cell! As discussed next, certain modifications to PCR can allow researchers to observe the accumulation of PCR products and to quantitate the amount of DNA or RNA in a biological sample.

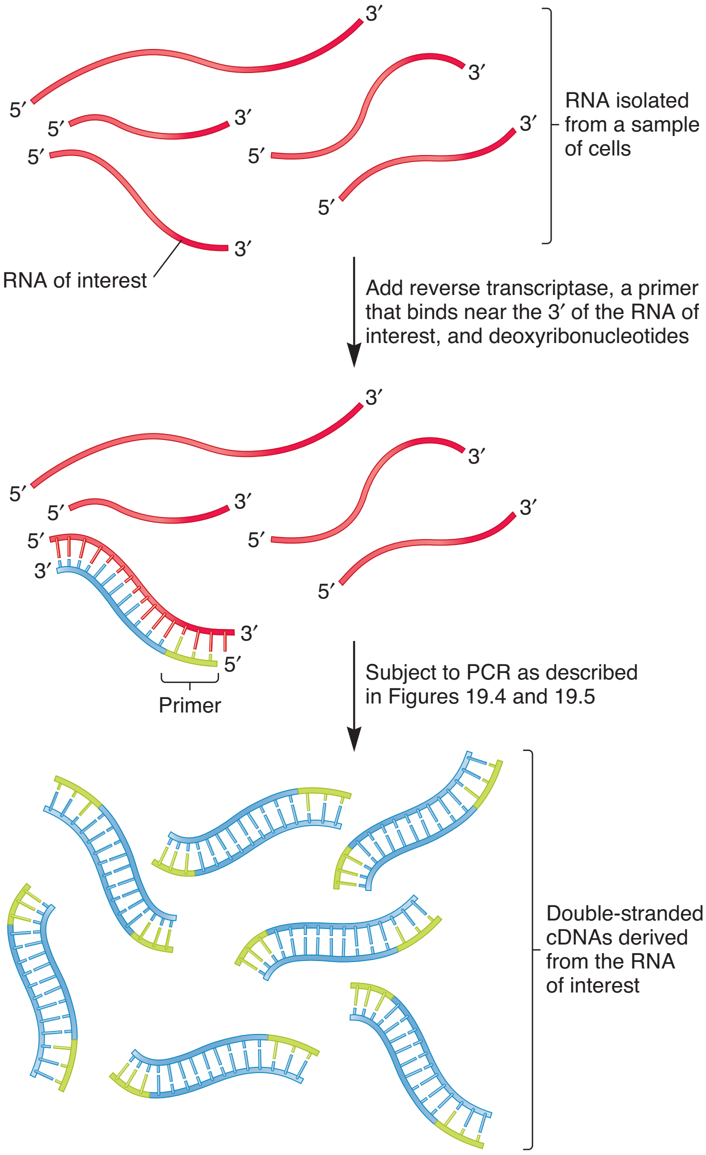

3′
5′
5′
3′
5′
3′
3′
5′
3′
5′
3′
5′
RNA isolated from a sample of cells
RNA of interest
Add reverse transcriptase, a primer that binds near the 3′ of the RNA of interest, and deoxyribonucleotides
3′
5′
5′
3′
5′
3′
3′
5′
3′
5′
3′
5′
Primer
Subject to PCR as described in Figures 19.4 and 19.5
Double-stranded cDNAs derived from the RNA of interest

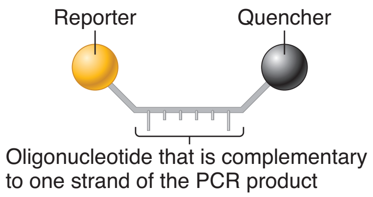
FIGURE 19.6 **The technique of reverse transcriptase PCR.**

Real-Time PCR Can Be Used to Quantitate the Amount of a Specific Gene or a Specific mRNA in a Sample

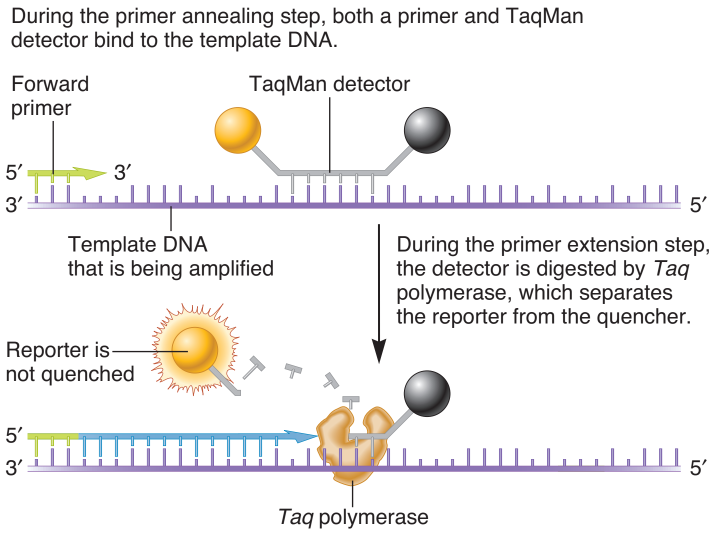
In some applications of PCR, the goal is to obtain large amounts of a DNA region. To determine if PCR is successful, a researcher typically runs a sample of DNA on a gel, stains the gel with ethidium bromide (EtBr) that binds to the DNA, and then observes the gel under UV light, which causes EtBr to fluoresce. (See the Appendix for a description of gel electrophoresis.) If a band of the correct size is seen, the experiment is likely to have been successful. This PCR approach is sometimes called endpoint analysis, because the success of the experiment is judged after PCR is completed.

By comparison, **real-time PCR** technology allows quantitation of specific PCR products in “real time” as PCR is taking place in a thermocycler. Because the PCR products are ultimately derived from the template DNA that was initially added to the reaction, researchers can use this approach to determine how much DNA (e.g., the DNA that encodes a specific gene) was originally in the sample before PCR was conducted. Similarly, if the starting material is mRNA that is reverse transcribed into DNA, real-time PCR can be used to determine how much mRNA from a specific gene was in a sample. This provides a way to quantitatively measure gene expression.

How do researchers determine the amount of a PCR product during real-time PCR? The procedure is carried out in a thermocycler that has the capacity to measure changes in the level of fluorescence that is emitted from detector molecules that are added to the PCR mixture. The fluorescence given off by the


Reporter
Quencher
Oligonucleotide that is complementary to one strand of the PCR product

(a) TaqMan detector


During the primer annealing step, both a primer and TaqMan detector bind to the template DNA.
Forward primer
TaqMan detector
5′
3′
3′
5′
Template DNA that is being amplified
During the primer extension step, the detector is digested by *Taq* polymerase, which separates the reporter from the quencher.
Reporter is not quenched
5′
3′
5′
Taq polymerase

(b) Use of a TaqMan detector in real-time PCR

FIGURE 19.7 An example of a detector, called TaqMan, which is used in a real-time PCR experiment.

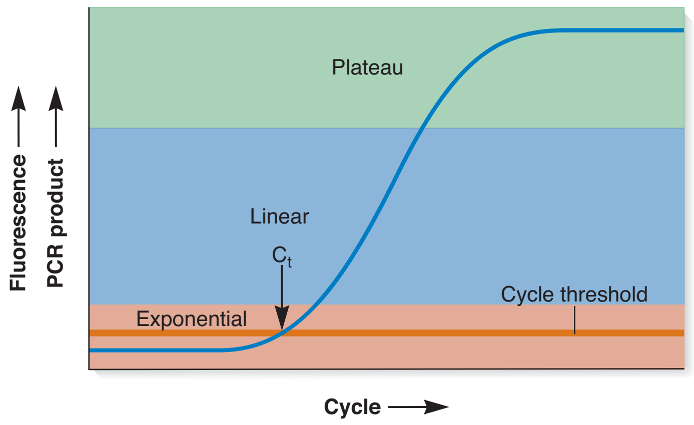
Concept check: What needs to happen so the reporter molecule can emit fluorescence that is not quenched?

detector molecules depends on the amount of the PCR product. A variety of detector molecules have been developed. We will consider one type called TaqMan.

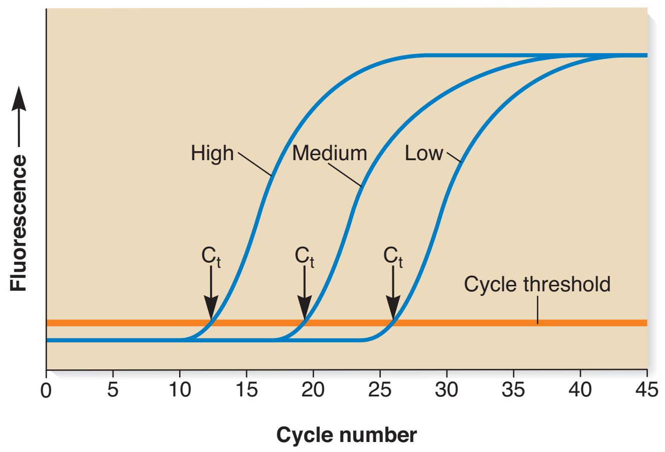
The TaqMan detector is an oligonucleotide that has a reporter molecule at one end and a quencher molecule at the other (**Figure 19.7a**). The oligonucleotide is complementary to a site within the PCR product of interest. The reporter molecule emits fluorescence at a certain wavelength, but that fluorescence is largely absorbed by the nearby quencher. Therefore, the close proximity of the reporter molecule to the quencher molecule prevents the detection of fluorescence from the reporter molecule.

After the denaturation step, a primer and TaqMan detector both anneal to the template DNA (**Figure 19.7b**). During the primer extension step, the 5′ to 3′ exonuclease activity of *Taq* polymerase cleaves the oligonucleotide in the TaqMan detector into individual nucleotides, thereby separating the reporter from the quencher. This allows the reporter to emit (unquenched) fluorescence that can be measured within the thermocycler. As PCR products accumulate, more and more of the TaqMan detectors are digested, and therefore the level of fluorescence increases.

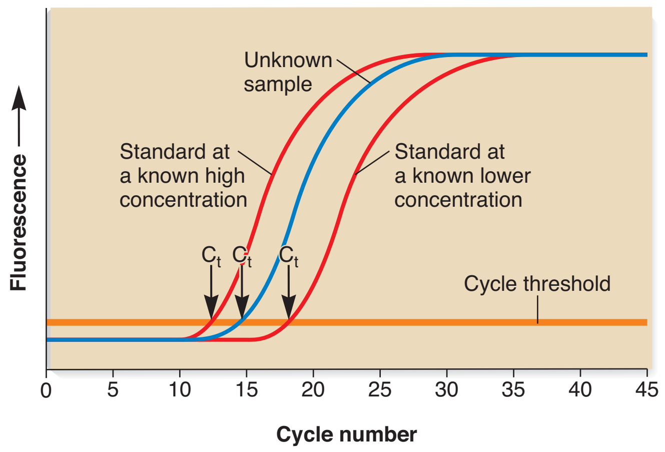
Figure 19.8a considers a real-time PCR experiment as it occurs over the course of many cycles, such as 20 to 40. Real-time PCR goes through three main phases. Initially, when the



(a) Phases of PCR



(b) Real-time PCR at high, medium, and low concentrations of the starting template DNA



(c) A comparison between an unknown sample and standards of known concentrations

FIGURE 19.8 Examples of data that are obtained from a real-time PCR experiment. (**a**) The three phases that occur in a typical PCR experiment. (**b**) PCR carried out at three different starting template concentrations. (**c**) A comparison between a sample with an unknown DNA concentration and a standard at a high and low concentration.

amount of PCR product is small and reagents are not limiting, the PCR product is made exponentially and the reaction is close to 100% efficiency. This exponential accumulation is difficult to detect in the earliest cycles because the amount of PCR product is small. Therefore, the amount of TaqMan detector that is cleaved is also small. The second phase is relatively linear as PCR products continue to accumulate, but the reaction efficiency falls as reagents diminish and become limiting. Finally, in the third phase, the accumulation of PCR products reaches a plateau as one or more reagents are used up.

During the exponential phase of PCR, the amount of PCR product is proportional to the amount of starting template that was initially added. Therefore, the exponential phase of PCR is analyzed to quantitate the initial template concentration. The commonly used method is called the **cycle threshold method (C_t method).** The cycle threshold (C_t) is reached when the accumulation of fluorescence is significantly greater than the background level. During the early cycles, the fluorescence signal due to the background level of fluorescence is greater than that derived from the amplification of the PCR product. Once the C_t value is exceeded, the exponential accumulation of product can be measured. **Figure 19.8b** considers three PCR runs in which the template DNA was initially added at a high, medium, or low concentration. When the initial concentration of the template DNA is higher, the C_t is reached at an earlier amplification cycle.

To determine the amount of starting template DNA, the sample of interest that has an unknown amount of starting template DNA is compared with some type of standard. Real-time PCR involves the coamplification of two templates: the sample of interest and the standard. The two types of PCR products can be monitored simultaneously with different-colored fluorescent molecules.

Two different types of standards may be used. One possibility is that a standard of known concentration is added to the PCR mixture. For example, plasmid DNA carrying a specific gene can be added to the PCR mixture in known amounts, and the amplification of this plasmid-encoded gene would provide a standard. By comparing the C_t values of the standard and the unknown sample, researchers can determine the concentration of the unknown sample. This is schematically shown in **Figure 19.8c**, but is actually done via computer software. Alternatively, researchers may use an internal standard in which another gene that is already present in the sample is also amplified. This relative quantitation method is somewhat simpler. The amount of unknown template DNA of interest is expressed relative to the internal standard.

19.2 REVIEWING THE KEY CONCEPTS

- Polymerase chain reaction (PCR) uses oligonucleotide primers to copy a specific region of DNA. Each cycle of PCR involves three steps: denaturation, primer annealing, and primer extension (see Figures 19.4, 19.5).
- Reverse transcriptase PCR is used to detect and quantitate the amount of an RNA of interest in living cells (see Figure 19.6).
- Real-time PCR monitors PCR as it occurs in a thermocycler. It can be used to quantitatively measure the amount of starting DNA or mRNA in a sample. Real-time PCR uses fluorescent detectors to follow the PCR reaction. A sample that has an unknown amount of DNA is compared with a standard (see Figures 19.7, 19.8).

19.2 COMPREHENSION QUESTIONS

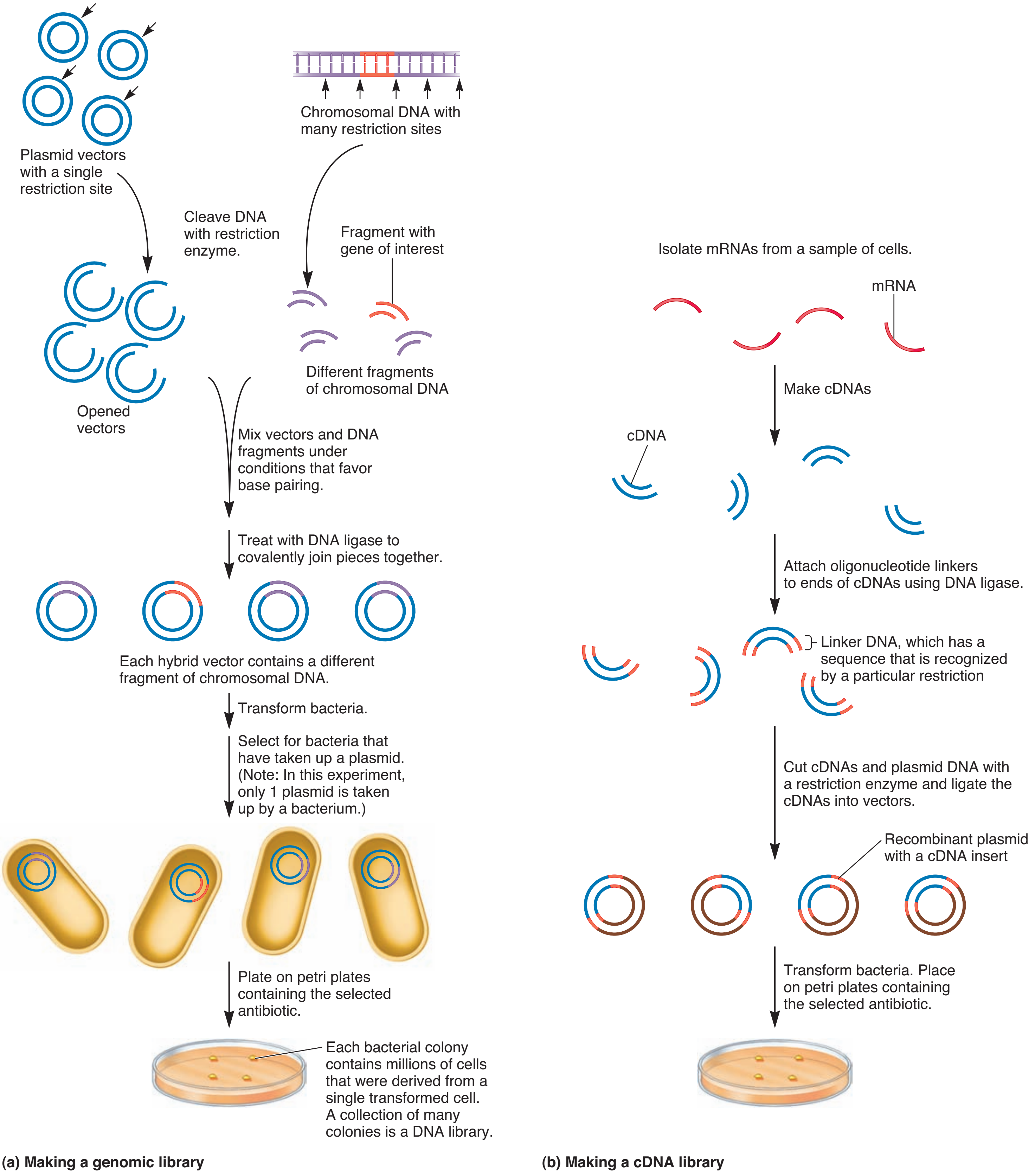
1. In one PCR cycle, the correct order of steps is
    a. primer annealing, primer extension, denaturation.
    b. primer annealing, denaturation, primer extension.
    c. denaturation, primer annealing, primer extension.
    d. denaturation, primer extension, primer annealing.
2. In reverse transcriptase PCR, the starting biological material is
    a. chromosomal DNA.
    b. mRNA.
    c. proteins.
    d. all of the above.
3. During real-time PCR, the synthesis of PCR products is analyzed
    a. at the very end of the reaction by gel electrophoresis.
    b. at the very end of the reaction by fluorescence that is emitted within the thermocycler.
    c. during the PCR cycles by gel electrophoresis.
    d. during the PCR cycles by fluorescence that is emitted within the thermocycler.

19.3 DNA LIBRARIES AND BLOTTING METHODS

The advent of gene cloning has enabled scientists to investigate gene structure and function at the molecular level. Molecular geneticists usually want to study genes that are originally within the chromosomes of living species. This presents a problem because chromosomal DNA carries thousands of different genes. For this reason, researchers need methods to specifically identify a gene within a mixture of many other genes or DNA fragments. Similarly, when studying gene expression at the molecular level, scientists also need techniques for the identification of gene products, such as the RNA that is transcribed from a particular gene or the protein that is encoded by an mRNA. In this section, we will consider the methodology and uses of several common techniques used to detect DNA, RNA, and proteins.

A DNA Library May Be Constructed Using Genomic DNA or cDNA

In a typical cloning experiment that involves the use of vectors (refer back to Figure 19.2), the treatment of the chromosomal DNA with restriction enzymes yields tens of thousands of different DNA fragments. Therefore, after the DNA fragments are ligated individually to vectors, the researcher has a collection of recombinant vectors, with each vector containing a particular fragment of chromosomal DNA. A collection of recombinant vectors is known as a **DNA library.** When the starting material is chromosomal DNA, the library is called a **genomic library** (**Figure 19.9a**). The library shown here uses a plasmid vector. Alternatively, a viral vector could be used, which would result in viral plaques rather than bacterial colonies.


Plasmid vectors with a single restriction site
Chromosomal DNA with many restriction sites
Cleave DNA with restriction enzyme.
Fragment with gene of interest
Opened vectors
Different fragments of chromosomal DNA
Mix vectors and DNA fragments under conditions that favor base pairing.
Treat with DNA ligase to covalently join pieces together.
Each hybrid vector contains a different fragment of chromosomal DNA.
Transform bacteria.
Select for bacteria that have taken up a plasmid. (Note: In this experiment, only 1 plasmid is taken up by a bacterium.)
Plate on petri plates containing the selected antibiotic.
Each bacterial colony contains millions of cells that were derived from a single transformed cell. A collection of many colonies is a DNA library.
Isolate mRNAs from a sample of cells.
mRNA
Make cDNAs
cDNA
Attach oligonucleotide linkers to ends of cDNAs using DNA ligase.
Linker DNA, which has a sequence that is recognized by a particular restriction
Cut cDNAs and plasmid DNA with a restriction enzyme and ligate the cDNAs into vectors.
Recombinant plasmid with a cDNA insert
Transform bacteria. Place on petri plates containing the selected antibiotic.

(a) Making a genomic library

(b) Making a cDNA library

FIGURE 19.9 **The construction of a DNA library.** (**a**) To make a genomic library, chromosomal DNA is first digested to produce many fragments. The fragment containing the gene of interest is highlighted in red. Following ligation, each vector contains a different piece of chromosomal DNA. (**b**) To make a cDNA library, oligonucleotide linkers that contain a restriction site are attached to the cDNAs, so they can be inserted into vectors.

Concept check: *What is an advantage of making a cDNA library rather than a genomic library?*

It is also common for researchers to make a **cDNA library** that contains recombinant vectors with cDNA inserts. Because cDNA is produced from mature mRNA via reverse transcriptase, it lacks introns. A cDNA library could be made because a researcher wanted to express the encoded protein of interest in a cell that would not splice out the introns properly. Figure 19.9b illustrates how cDNAs are made and inserted into vectors. As described earlier in Figure 19.3, cDNA is first made via reverse transcriptase. To insert the cDNAs into vectors, short oligonucleotides called linkers are attached to the cDNAs via DNA ligase. The linkers contain DNA sequences with a unique site for a restriction enzyme. After the linkers are attached to the cDNAs, the cDNAs and the vectors are cut with restriction enzymes and then ligated to each other. This produces a cDNA library.

A DNA Library May Be Screened by Colony Hybridization to Identify a Cloned Gene

In many cases, the goal of a gene-cloning experiment is to obtain many copies of a specific gene or region of a gene. In cloning via PCR, this is achieved by selecting primers that flank the region of interest. This allows a researcher to specifically clone a gene of interest without cloning other segments of chromosomal DNA. However, when cloning is done using vectors, many different segments of chromosomal DNA are cloned during the construction of a DNA library. For example, let's suppose that a geneticist wishes to clone the rat β-globin gene. To begin a cloning experiment, chromosomal DNA is isolated from rat cells. This chromosomal DNA is then digested with a restriction enzyme, yielding thousands of DNA fragments. The chromosomal fragments are then ligated to vector DNA and transformed into bacterial cells to make a DNA library. Unfortunately, only a small percentage of the recombinant vectors, perhaps one in a few thousand, actually contains the rat β-globin gene. For this reason, researchers must have some way of identifying those rare bacterial colonies within a DNA library that happen to contain the cloned gene of interest, in this case, a colony that contains the rat β-globin gene.

Figure 19.10 describes one method for identifying a bacterial colony that contains the gene of interest. This procedure is referred to as **colony hybridization.** The master plate shown at the top of Figure 19.10 has many bacterial colonies. Each bacterial colony is composed of bacterial cells containing a recombinant vector with a different piece of rat chromosomal DNA. The goal is to identify a colony that contains the gene of interest, in this case, the rat β-globin gene. To do so, a nylon membrane is laid gently onto the master plate containing many bacterial colonies. After the membrane is lifted, some cells from each colony are attached to it. In this way, the membrane contains a replica of the colonies on the master plate.

In this procedure, cells are treated with detergent, which dissolves the cell membrane and makes the DNA accessible. The DNA within the cells is then fixed (i.e., adhered) to the nylon membrane and denatured with NaOH. The membrane is then submerged in a solution containing a **probe,** which can be a oligonucleotide, a fragment of DNA from a cloned gene, or a specific RNA. In this case, the probe is single-stranded DNA with a base sequence that is complementary to one of the DNA strands of the rat β-globin gene. The probe is labeled in some way for ease of detection. It could be radiolabeled using a radioisotope such as ^{32}P, or it could be fluorescently labeled. The example in Figure 19.10 uses a radiolabeled probe. The probe is given time to hybridize to the DNA on the membrane, which has already been denatured into single strands. If a bacterial colony contains the rat β-globin gene, the probe will hybridize to the DNA in this colony. Most bacterial colonies are not expected to contain this gene. The unbound probe is then washed away, and the membrane is placed next to X-ray film. If the DNA within a bacterial colony did hybridize to

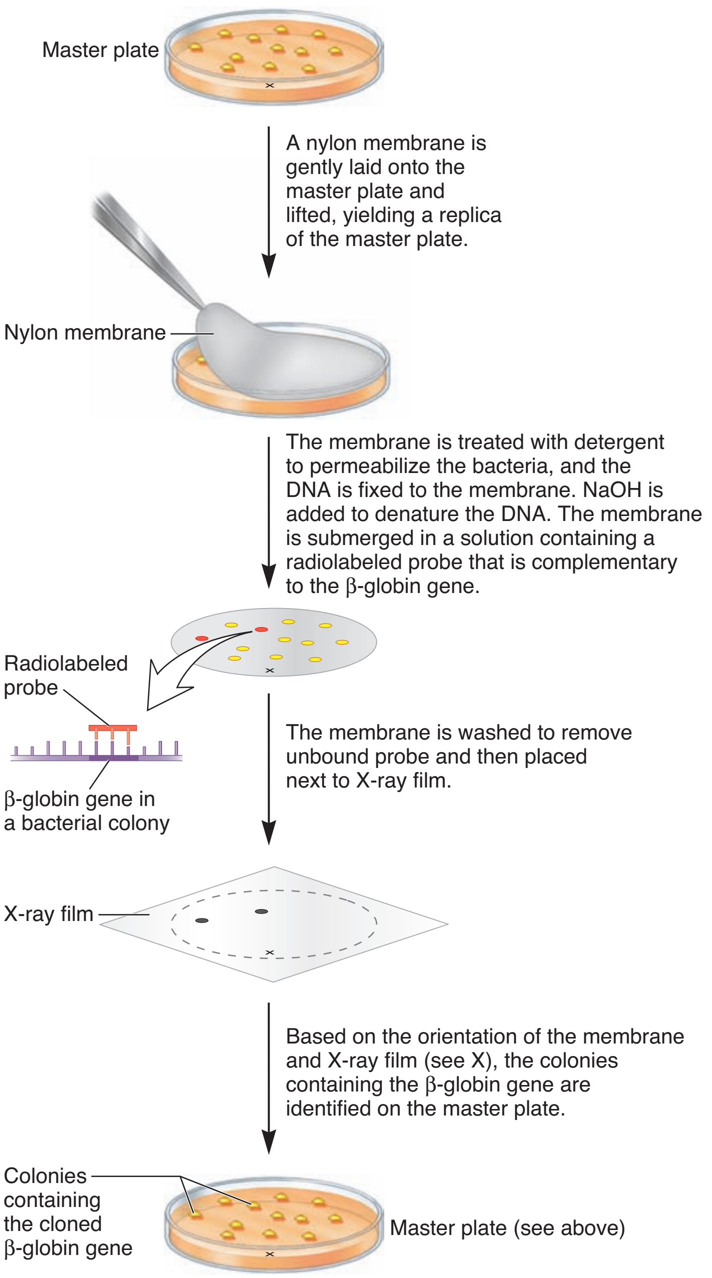

Master plate
A nylon membrane is gently laid onto the master plate and lifted, yielding a replica of the master plate.
Nylon membrane
The membrane is treated with detergent to permeabilize the bacteria, and the DNA is fixed to the membrane. NaOH is added to denature the DNA. The membrane is submerged in a solution containing a radiolabeled probe that is complementary to the β-globin gene.
Radiolabeled probe
β-globin gene in a bacterial colony
The membrane is washed to remove unbound probe and then placed next to X-ray film.
X-ray film
Based on the orientation of the membrane and X-ray film (see X), the colonies containing the β-globin gene are identified on the master plate.
Colonies containing the cloned β-globin gene
Master plate (see above)

FIGURE 19.10 **A colony hybridization experiment.** Note: If the vector used was a viral vector, the master plate would have viral plaques. The steps shown here could also be followed, in which case the experiment would be called plaque hybridization.

the probe, a dark spot will appear on the film in the corresponding location. The developed film is compared with the master plate to identify bacterial colonies carrying the rat β-globin gene. Following the identification of labeled colonies, the experimenter can pick them from the plate, and grow bacteria containing the cloned gene.

A scientist might follow several different strategies to obtain a probe for a colony hybridization experiment. In some cases, the gene of interest may have already been cloned. A piece of the cloned gene then can be used as a radiolabeled probe. However, in many circumstances, a scientist may attempt to clone a gene that has never been cloned before. In such a case, one strategy is to use a probe that likely has a sequence similar to that of the gene of interest. For example, if the goal is to clone the β-globin gene from a South American rodent, it is expected that this gene is similar to the rat β-globin gene, which has already been cloned. Therefore, the cloned rat β-globin gene can be used as a DNA probe to "fish out" the β-globin gene from a similar species.

Can a researcher design a DNA probe for a novel type of gene that no one else has ever cloned before from any species? If the protein of interest has been previously isolated from living cells, fragments of the protein may be subjected to amino acid sequence analysis to obtain short amino acid sequences. Because the genetic code is known, a researcher could use this amino acid sequence information to design oligonucleotide probes that correspond to the coding sequence of the gene that encodes the protein. Alternatively, the purified protein could be used to generate antibodies. In that case, the antibody would recognize clones that express the protein. Like DNA probes, labeled antibodies can be used as probes in a colony hybridization experiment. The use of antibodies as labeling tools in the technique known as Western blotting will be discussed later in this chapter.

Southern Blotting Is Used to Detect DNA Sequences

The technique of **Southern blotting** is used to detect the presence of a particular gene sequence within a mixture of many chromosomal DNA fragments. This method, developed by Edwin Southern in 1975, has several uses. It can determine the copy number of a gene within the genome of an organism. For example, Southern blotting has revealed that rRNA genes are found in multiple copies within a genome, whereas many structural genes are unique. In the study of human genetic diseases, Southern blotting can also detect small gene deletions that cannot be distinguished under the light microscope (see Solved Problem S4 at the end of this chapter). A common use of Southern blotting is to identify gene families. As discussed in Chapter 8, a gene family is a group of two or more genes derived from the same ancestral gene. The members of a gene family have similar but not identical DNA sequences; they are homologous genes. As we will see, Southern blotting can distinguish the homologous members of a gene family. Similarly, Southern blotting can identify homologous genes in different species. This is called a zoo blot. Finally, as discussed in Chapter 20, researchers may produce transgenic organisms that carry new or modified genes. Southern blotting may be used to verify that transgenic organisms carry such genes.

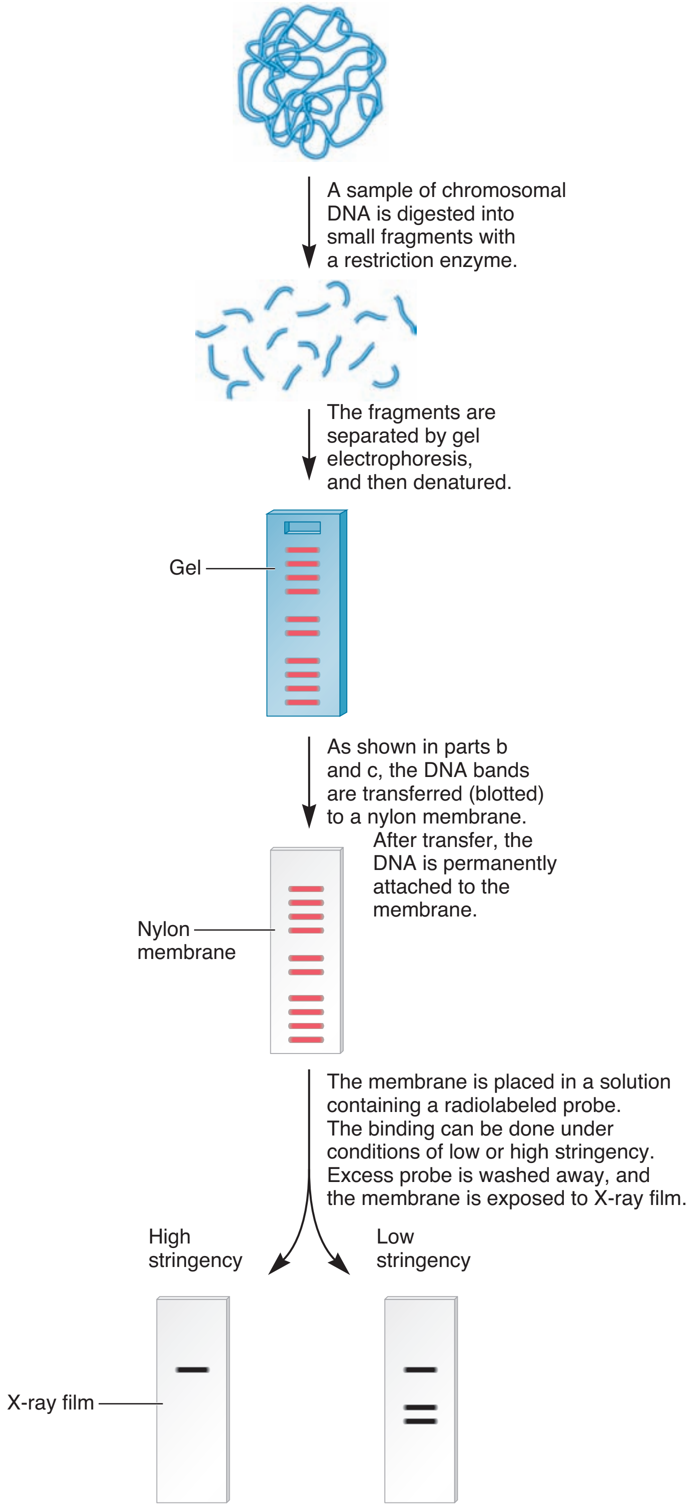
Prior to a Southern blotting experiment, a gene or fragment of a gene of interest has been cloned. This cloned DNA is labeled (e.g., radiolabeled) in vitro, and then the labeled DNA is used as a probe to detect the presence of the gene or a homologous gene within a mixture of many DNA fragments. The basis for a Southern blotting experiment is that two DNA fragments bind to each other only if they have complementary sequences. In such an experiment, the labeled strands from the cloned gene pair specifically with complementary DNA strands, even if the complementary strands are found within a mixture of many other DNA pieces.

Figure 19.11a shows the Southern blotting procedure. The goal of this experiment is to determine if the chromosomal DNA contains a base sequence complementary to a specific probe. To begin the experiment, the chromosomal DNA is isolated and digested with a restriction enzyme. Because the restriction enzyme cuts the chromosomal DNA at many different sites within the chromosomes, this step produces thousands of DNA pieces of different sizes. Depending on the location of the restriction sites, each gene is contained on one or more DNA pieces with specific masses. The chromosomal pieces are separated according to their masses by gel electrophoresis. The DNA pieces within the gel are denatured by soaking the gel in a NaOH solution, and then the DNA is transferred (i.e., blotted) to a nylon membrane.

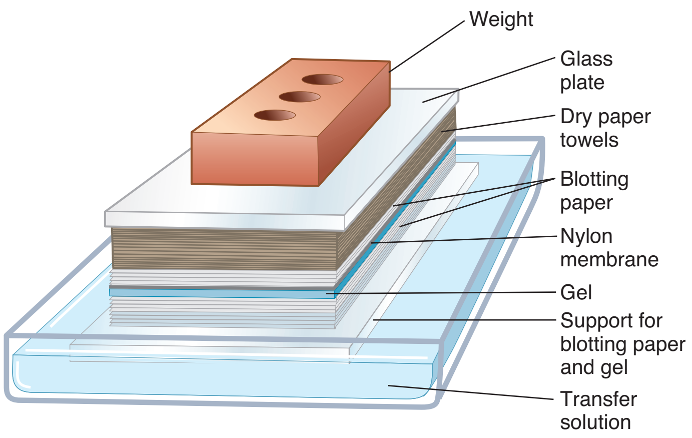
As shown in **Figure 19.11b**, the traditional way to transfer the DNA to the membrane is to lay the gel on blotting paper in a transfer solution. A nylon membrane is placed on top of the gel, and then additional blotting paper and dry paper towels are layered above the nylon membrane. A glass plate and heavy weight are placed on the very top. The liquid from the transfer solution is drawn upward, carrying the DNA from the gel to the nylon membrane. When the DNA enters the nylon membrane, it remains bound to it. Alternatively, another way to transfer DNA is via a transfer apparatus in which the gel is placed on top of a nylon membrane and then an electric field causes the DNA fragments to electrophoretically move from the gel into the membrane (**Figure 19.11c**).

The procedures shown in Figure 19.11b and c result in a nylon membrane that contains many unlabeled DNA fragments that have been separated according to mass and then denatured. The nylon membrane is incubated in an oven or exposed to UV light to permanently attach the DNA fragments to the membrane. The next step is to determine if any of these unlabeled fragments from the chromosomal DNA contain sequences that are complementary to a labeled probe. One labeling method is to incorporate the radioisotope ^{32}P into the probe DNA, which labels the phosphate group in the DNA backbone. The nylon membrane, which has the unlabeled chromosomal DNA attached to it, is submerged in a solution containing the radiolabeled probe. If the radiolabeled probe and a fragment of chromosomal DNA are complementary, they will hydrogen bond to each other. Any unbound radiolabeled DNA is then washed away, and the membrane is exposed to X-ray film. Locations where the radiolabeled probe has bound appear as dark bands on the X-ray film.

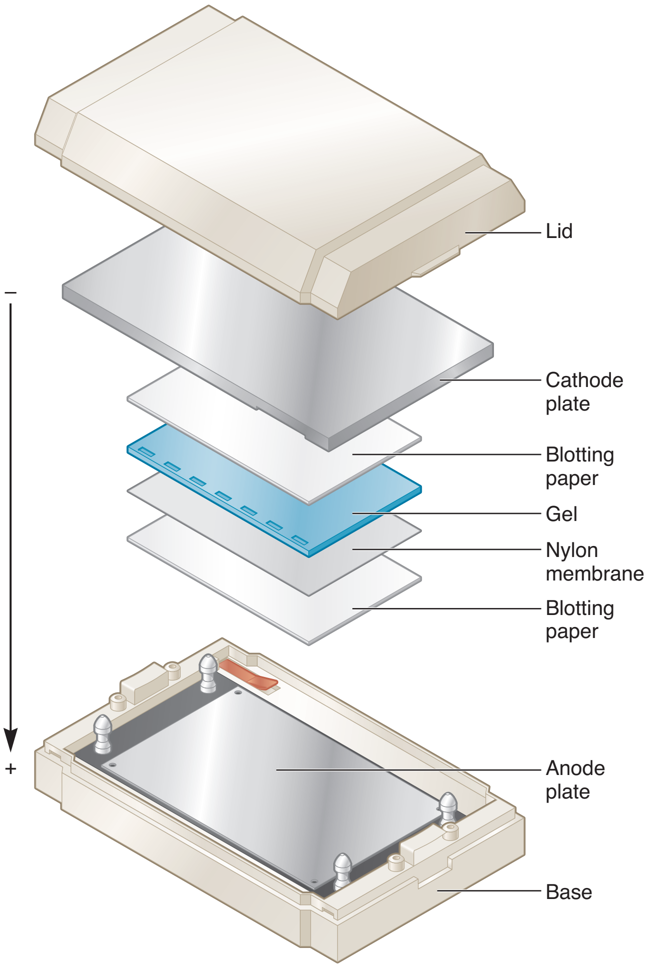
Important variables in the Southern blotting procedure are the temperature and ionic strength of the hybridization and wash steps. If the procedure is done at very high temperatures


A sample of chromosomal DNA is digested into small fragments with a restriction enzyme.
The fragments are separated by gel electrophoresis, and then denatured.
Gel
As shown in parts b and c, the DNA bands are transferred (blotted) to a nylon membrane.
After transfer, the DNA is permanently attached to the membrane.
Nylon membrane
The membrane is placed in a solution containing a radiolabeled probe. The binding can be done under conditions of low or high stringency. Excess probe is washed away, and the membrane is exposed to X-ray film.
High stringency
Low stringency
X-ray film

(a) The steps in Southern blotting


Weight
Glass plate
Dry paper towels
Blotting paper
Nylon membrane
Gel
Support for blotting paper and gel
Transfer solution

(b) Transfer step (traditional method)


Lid
–
Cathode plate
Blotting paper
Gel
Nylon membrane
Blotting paper
+
Anode plate
Base

(c) The transfer step via electrophoresis

FIGURE 19.11 **The technique of Southern blotting.** **(a)** Chromosomal DNA is isolated from a sample of cells and then digested with a restriction enzyme to yield many DNA fragments. The fragments are then separated by gel electrophoresis according to molecular mass and denatured with NaOH. The bands within the gel are transferred to a nylon membrane, a step known as blotting. The blot is then placed in a solution containing a labeled DNA probe. When the DNA probe is hybridized to the mixture of chromosomal DNA fragments at high temperature and/or low ionic strength (i.e., high stringency), the probe recognizes only a single band. When hybridization is performed at lower temperature and/or higher ionic strength (i.e., low stringency), this band and two additional bands are produced. **(b)** The transfer step of DNA blotting (traditional method). **(c)** The transfer step of DNA blotting via electrophoresis.

Concept check: *What type of information is obtained under conditions of low stringency?*

and/or at low salt concentrations, the probe DNA and chromosomal fragment must be very complementary—nearly a perfect match—in order to hybridize. This condition is called **high stringency.** Conditions of high stringency are used to detect close or exact complementarity between the probe and a chromosomal DNA fragment. However, if the temperature is lower and/or the salt concentration is higher, DNA sequences that are not perfectly complementary may hybridize to the probe. This is called **low stringency.** Conditions of low stringency are used to detect homologous genes with DNA sequences that are similar but not identical to the cloned gene being used as a probe. In the results shown in Figure 19.11a, conditions of high stringency reveal that the gene of interest is unique in the genome. At low stringency, however, two other bands are detected. These results suggest that this gene is a member of a gene family composed of three distinct members.

Northern Blotting Is Used to Detect RNA

As discussed earlier in this chapter, reverse transcriptase PCR and real-time PCR are used to detect and quantitate the amount of RNA in a biological sample. Another approach, known as **Northern blotting,** can also be used to identify a specific RNA within a mixture of many RNA molecules. (Note: Even though Southern blotting was named after Edwin Southern, Northern blotting was not named after anyone named Northern! It was originally termed reverse-Southern blotting and later became known as Northern blotting.) Northern blotting is used to investigate the transcription of genes at the molecular level. This method can determine if a specific gene is transcribed in a particular cell type, such as nerve or muscle cells, or at a particular stage of development, such as fetal or adult cells. Also, Northern blotting can reveal if a pre-mRNA transcript is alternatively spliced into two or more mRNAs of different sizes.

From a technical viewpoint, Northern blotting is similar to Southern blotting with a few important differences. RNA is the starting material and is extracted and purified from living cells. This RNA can be isolated from a particular cell type under a given set of conditions or during a particular stage of development. Any given cell produces thousands of different types of RNA molecules, because cells express many genes at any given time. After the RNA is extracted from cells and purified, it is loaded onto a gel that separates the RNA transcripts according to their size. The RNAs within the gel are then blotted onto a nylon membrane and probed with a labeled fragment of DNA from a cloned gene. If the probe is radiolabeled, RNAs that are complementary to the probe DNA are detected as dark bands on an X-ray film.

In Figure 19.12, the probe was complementary to an mRNA that encodes a protein called tropomyosin. In lane 1, the RNA was isolated from smooth muscle cells, in lane 2 from striated muscle cells, and in lane 3 from brain cells. As seen here, smooth and striated muscle cells contain a large amount of this mRNA. This result is expected because tropomyosin plays a role in the regulation of cell contraction. By comparison, brain cells have much less of this mRNA. In addition, we see that the molecular masses of the three mRNAs are different among the three cell types. This observation indicates that the pre-mRNA is alternatively spliced to contain different combinations of exons.

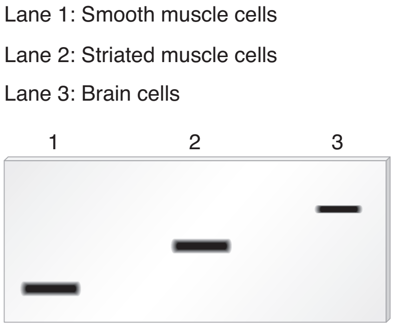


FIGURE 19.12 **The results of Northern blotting.** RNA is isolated from a sample of cells and then separated by gel electrophoresis. The separated RNA bands are blotted to a nylon membrane and then placed in a solution containing a radioactive probe. Following autoradiography, RNA molecules that are complementary to the probe appear as dark bands on the X-ray film.

Western Blotting Is Used to Detect Proteins

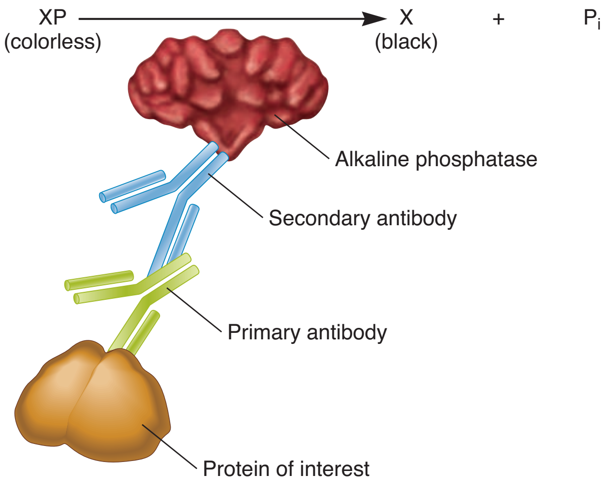
For structural genes, the end result of gene expression is the synthesis of proteins. A particular protein within a mixture of many different protein molecules can be identified by **Western blotting,** a third detection procedure. This process, like Northern blotting, was named as part of a group of blotting techniques, not for anyone named Western. This method can determine if a specific protein is made in a particular cell type or at a particular stage of development. Technically, this procedure is also similar to Southern and Northern blotting.

In a Western blotting experiment, proteins are extracted from living cells. As with RNA, any given cell produces many different proteins at any time, because it is expressing many structural genes. After the proteins have been extracted from the cells, they are loaded onto a gel that separates them by molecular mass. To perform the separation step, the proteins are first dissolved in sodium dodecyl sulfate (SDS), a detergent that denatures proteins and coats them with negative charges. The negatively charged proteins are then separated in a gel made of polyacrylamide. This method of separating proteins is called SDS-PAGE (polyacrylamide gel electrophoresis).

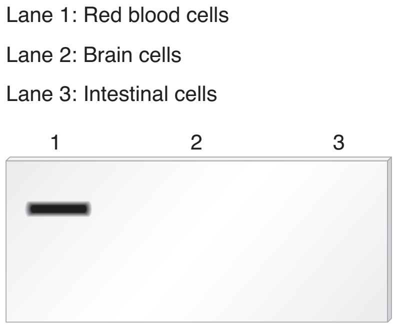
Following SDS-PAGE, the proteins within the gel are blotted to a nylon membrane. The next step is to use a probe that recognizes a specific protein of interest. An important difference between Western blotting and either Southern or Northern blotting is the use of an **antibody** as a probe, rather than a labeled DNA strand. Antibodies bind to sites known as **epitopes.** An epitope has a three-dimensional structure that is recognized by an antibody. The term **antigen** refers to any molecule that is recognized by an antibody. An antigen contains one or more epitopes. In the case of proteins, an epitope is a short sequence of amino acids. Because the amino acid sequence is a unique feature of each protein, any given antibody specifically recognizes a particular protein. In a Western blot, this is called the primary antibody for that protein.

After the primary antibody has been given sufficient time to recognize the protein of interest, any unbound primary antibody is washed away, and a secondary antibody is added. A secondary antibody is an antibody that binds to the primary antibody. Secondary antibodies, which may be radiolabeled or conjugated to an enzyme, are used for convenience, because secondary antibodies are available commercially. In general, it is easier for researchers to obtain these antibodies from commercial sources rather than labeling their own primary antibodies. In a Western blotting experiment, the secondary antibody provides a way to detect the protein of interest in a gel blot. For example, it is common for the secondary antibody to be linked to the enzyme alkaline phosphatase. When the colorless compound XP (5-bromo-4-chloro-3-indolyl-phosphate) is added to the blotting solution, alkaline phosphatase converts the compound to a black dye (**Figure 19.13a**). Because the secondary antibody binds to the primary antibody, a protein band that is recognized by the primary antibody becomes black.

In the Western blot shown in **Figure 19.13b**, the primary antibody recognized β globin. In lane 1, proteins were isolated from mouse red blood cells. The black band indicates that the β-globin polypeptide is made in these cells. By comparison, the proteins loaded into lanes 2 and 3 were from brain cells and intestinal cells, respectively. The absence of any bands indicates that these cell types do not synthesize β globin.



(a) Interactions between the protein of interest and antibodies



(b) Results from a Western blotting experiment

FIGURE 19.13 **Western blotting. (a)** To detect a specific protein, a primary antibody first binds to the protein of interest and then a secondary antibody binds to the primary antibody. In this example, the secondary antibody is also attached to an enzyme called alkaline phosphatase that converts XP to a black color. **(b)** The black band indicates where the primary antibody has recognized the protein of interest. In lane 1, proteins were isolated from mouse red blood cells. As seen here, the β-globin polypeptide is made in these cells. By comparison, lanes 2 and 3 were samples of protein from brain cells and intestinal cells, which do not synthesize β globin.

Concept check: What is the purpose of using a secondary antibody?

19.3 REVIEWING THE KEY CONCEPTS

- A DNA library is a collection of recombinant vectors. When the starting material is chromosomal DNA, the library is called a genomic library; when the starting material is mRNA, it is called a cDNA library (see Figure 19.9).
- Colony hybridization is used to identify colonies that carry a gene or cDNA of interest (see Figure 19.10).
- Southern blotting uses a labeled DNA probe to detect the presence of a particular gene sequence within a mixture of many gene sequences. At low stringency, gene families may be detected (see Figure 19.11).
- Northern blotting uses a labeled DNA probe to detect a specific RNA within a mixture of many different RNAs (see Figure 19.12).
- Western blotting uses antibodies to detect a specific protein within a mixture of many different proteins (see Figure 19.13).

19.3 COMPREHENSION QUESTIONS

1. A collection of recombinant vectors that contain fragments of chromosomal DNA is called
   a. a genomic library.
   b. a cDNA library.
   c. a Southern blot.
   d. both a and b.
2. Which of the following methods is a hybridization method that uses a labeled nucleic acid probe, such as a radiolabeled fragment of DNA?
   a. Southern blotting
   b. Northern blotting
   c. Western blotting
   d. Both a and b
3. Which of the following methods is used to detect a specific RNA within a mixture of many different RNAs?
   a. Southern blotting
   b. Northern blotting
   c. Western blotting
   d. None of the above
4. During Western blotting, the primary antibody recognizes
   a. the secondary antibody.
   b. the protein of interest.
   c. an mRNA of interest.
   d. a specific fragment of chromosomal DNA.

19.4 METHODS FOR ANALYZING DNA- AND RNA-BINDING PROTEINS

In addition to detecting the presence of genes and gene products using blotting techniques, researchers often want to study the binding of proteins to specific sites on a DNA or RNA molecule. For example, the molecular investigation of transcription factors requires methods for identifying interactions between transcription factor proteins and specific DNA sequences. In Chapter 22, we will consider the technique of chromatin immunoprecipitation, which can determine whether proteins bind to a particular region of DNA found within the chromatin of living cells (look ahead to Figure 22.2). In this section, we will examine two additional methods: the gel retardation assay and DNase I footprinting, which are used to study protein-DNA interactions. Gel retardation is also used to study protein-RNA interactions.

The Gel Retardation Assay Is Used to Determine If a Protein Binds to a Specific DNA Fragment or RNA Molecule

A technically simple, widely used method for identifying DNA- or RNA-binding proteins is the **gel retardation assay,** also known as the **gel mobility shift assay.** This technique was used originally to study interactions between specific proteins and rRNA molecules, and quickly became popular after its success in studying protein-DNA interactions in the *lac* operon. Now it is commonly used as a technique for detecting interactions between eukaryotic transcription factors and DNA regulatory elements and between mRNAs and RNA-binding proteins.

In the case of DNA-binding proteins, the technical basis for a gel retardation assay is that the binding of a protein to a DNA fragment retards the DNA's ability to move within a polyacrylamide or agarose gel. During electrophoresis, DNA fragments are pulled through the gel matrix toward the bottom of the gel by a voltage gradient. Smaller fragments of DNA migrate more quickly through a gel matrix than do larger ones. The protein-DNA complex has a higher mass and migrates more slowly to the bottom of the gel. When comparing a DNA fragment and a protein-DNA complex after electrophoresis, the complex is shifted to a higher band compared to the DNA alone (**Figure 19.14**). The bands are visualized by staining the DNA with a dye such as EtBr. To increase the sensitivity of the gel retardation assay, the DNA can also be radiolabeled or labeled with a fluorescent molecule.

A gel retardation assay must be carried out under nondenaturing conditions. This means that the buffers and gel cannot cause the unfolding of proteins or the separation of the DNA double helix. This is necessary so the proteins and DNA retain their proper structure and are thereby able to bind to each other. The nondenaturing conditions of a gel retardation assay differ from the more common SDS-PAGE method of separating proteins, in which the proteins are denatured by the detergent SDS.

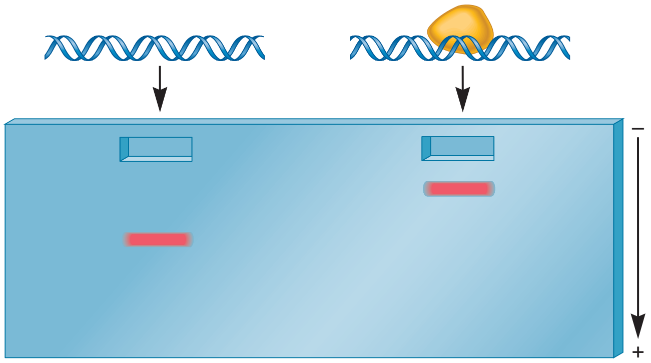


FIGURE 19.14 **The results of a gel retardation assay.** The binding of a protein to a labeled fragment of DNA retards its rate of movement through a gel. For the results shown in the lane on the right, if the concentration of the DNA fragment was higher than the concentration of the protein, there would be two bands: one band with protein bound (at a higher molecular mass) and one band without protein bound (corresponding to the band found in the left lane).

DNase I Footprinting Shows the Detailed Interactions Between a Protein and DNA

Another method for studying protein-DNA interactions is **DNase I footprinting,** a technique described by David Galas and Albert Schmitz in 1978. A DNase I footprinting experiment attempts to identify one or more regions of DNA that interact with a DNA-binding protein. In their original study, Galas and Schmitz identified a site in the *lac* operon, known as the *lac* operator site, that is bound by a DNA-binding protein called the lac repressor. Compared with a gel retardation assay, DNase I footprinting provides more detailed information about the interactions between a protein and DNA. One drawback is that DNase I footprinting is more technically difficult than a gel retardation assay.

To understand the basis of a DNase I footprinting experiment, we need to consider the interactions among three types of molecules: a fragment of DNA, DNA-binding proteins, and agents that alter DNA structure. As an example, let's examine the binding of RNA polymerase to a bacterial promoter, a topic discussed in Chapter 14. When RNA polymerase holoenzyme binds to the promoter to form a closed complex, it binds tightly to the −35 and −10 promoter regions, but the protein covers up an even larger region of the DNA. Therefore, holoenzyme bound at the promoter prevents other molecules from gaining access to this region of the DNA. The enzyme DNase I, which can cleave covalent bonds in the DNA backbone, is used as a reagent for determining if a DNA region has a protein bound to it. Galas and Schmitz reasoned that DNase I cannot cleave the DNA at locations where a protein is bound. In this example, it is expected that RNA polymerase holoenzyme will bind to a promoter and protect this DNA region from DNase I cleavage.

Figure 19.15 shows the results of a DNase I footprinting experiment. In this experiment, a sample of many identical DNA fragments, all of which are 150 bp in length, were radiolabeled at only one end. The sample of fragments was then divided into two

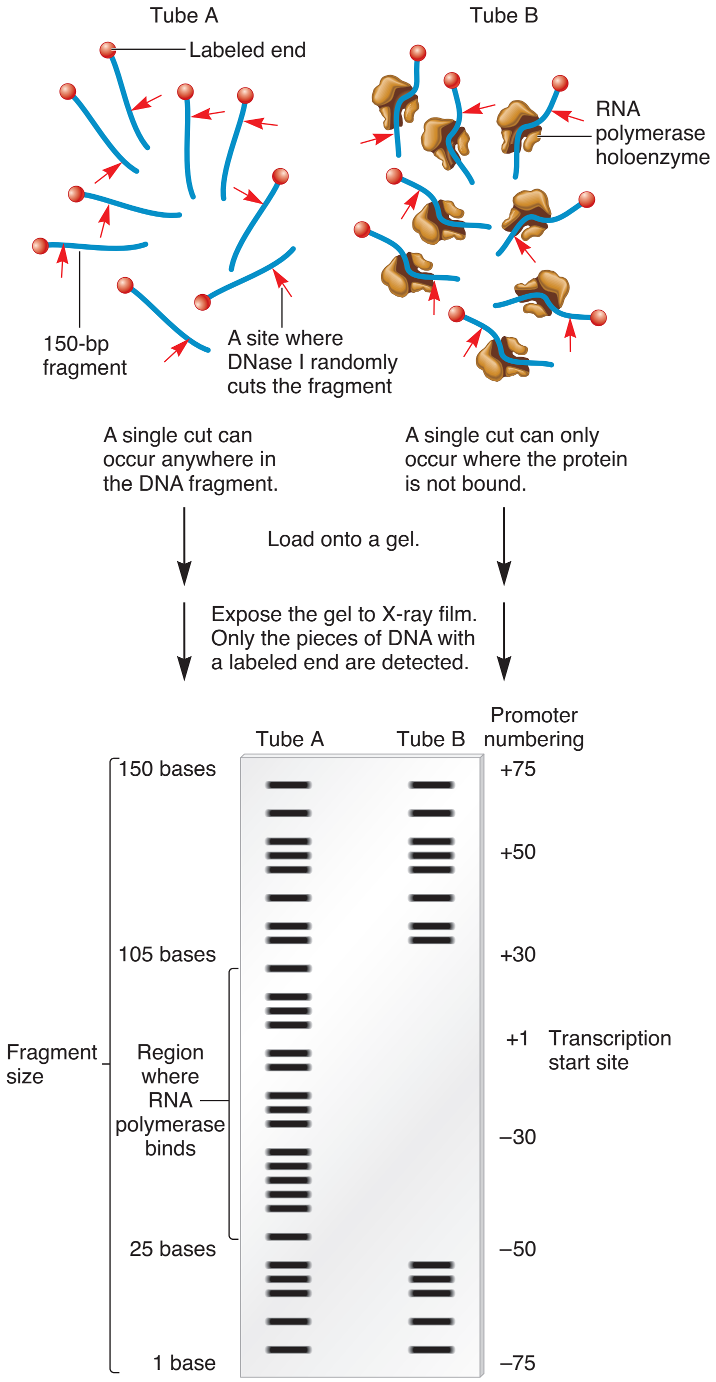

Tube A
Tube B
Labeled end
RNA polymerase holoenzyme
150-bp fragment
A site where DNase I randomly cuts the fragment
A single cut can occur anywhere in the DNA fragment.
A single cut can only occur where the protein is not bound.
Load onto a gel.
Expose the gel to X-ray film. Only the pieces of DNA with a labeled end are detected.
Tube A
Tube B
Promoter numbering
150 bases
105 bases
25 bases
1 base
Fragment size
Region where RNA polymerase binds
+75
+50
+30
+1
Transcription start site
−30
−50
−75

FIGURE 19.15 **A DNase I footprinting experiment.** Both tubes contained 150-bp fragments of DNA that were incubated with DNase I. Tube B also contained RNA polymerase holoenzyme. The binding of RNA polymerase holoenzyme protected a region of about 80 bp (namely, the −50 region to the +30 region) from DNase I digestion. Note: The promoter numbering convention shown here is the same as that described previously in Figure 14.4.

Concept check: *How does the binding of a protein to DNA influence the ability of DNase I to cleave the DNA?*

tubes: tube A, which did not contain any holoenzyme, and tube B, which contained RNA polymerase holoenzyme. DNase I was then added to both tubes. The tubes were incubated long enough for DNase I to cleave the DNA at a single site in each DNA fragment. Each tube contained many 150-bp DNA fragments, and the cutting in any DNA strand by DNase I occurred randomly. Therefore, the DNase I treatment produced a mixture of many smaller DNA fragments. A key point, however, is that DNase I cannot cleave the DNA in a region where RNA polymerase holoenzyme is bound. After DNase I treatment, the DNA fragments within the two tubes were separated by gel electrophoresis, and DNA fragments containing the radiolabeled end were detected using autoradiography.

In the absence of RNA polymerase holoenzyme (tube A), DNase I cleaves the 150-bp fragments randomly at any single location. Therefore, a continuous range of sizes occurs (see Figure 19.15). However, if we look at the gel lane from tube B, no bands are observed in the size range from 25 to 105 nucleotides. Why are these bands missing? The answer is that DNase I cannot cleave the DNA within the region where the holoenzyme is bound. The middle portion of the 150-bp fragment contains a promoter sequence that binds the RNA polymerase holoenzyme. Along the right side of the gel, the bases are numbered according to their position within the gene. (The site labeled +1 is where transcription begins.) As seen here, RNA polymerase covers up a fairly large region (its “footprint”) of about 80 bp, from the −50 region to the +30 region.

As illustrated in this experiment, DNase I footprinting can identify the DNA region that interacts with a DNA-binding protein. In addition to RNA polymerase-promoter binding, DNase I footprinting has been used to identify the binding sites for many other types of DNA-binding proteins, such as eukaryotic transcription factors and histones. This technique has greatly facilitated our understanding of protein-DNA interactions.

19.4 REVIEWING THE KEY CONCEPTS

- A gel retardation assay can determine if a protein binds to a specific DNA fragment or RNA molecule because the binding of the protein slows down the movement of the DNA or RNA through a gel (see Figure 19.14).
- DNase I footprinting can determine the regions of a DNA molecule that are bound by a protein (see Figure 19.15).

19.4 COMPREHENSION QUESTIONS

1. In a gel retardation assay, the binding of a protein to DNA
   a. prevents the DNA from being digested with a restriction enzyme.
   b. causes the DNA to migrate more slowly through a gel.
   c. causes the DNA to migrate more quickly through a gel.
   d. inhibits the expression of any genes within the DNA.
2. The basis for DNase I footprinting is that the binding of a protein to DNA
   a. prevents the DNA from being digested with a restriction enzyme.
   b. enhances the ability of the DNA to be digested with a restriction enzyme.
   c. prevents the DNA from being digested with DNase I.
   d. enhances the ability of the DNA to be digested by DNase I.

19.5 DNA SEQUENCING AND SITE-DIRECTED MUTAGENESIS

As we have seen throughout this textbook, our knowledge of genetics can be largely attributed to an understanding of DNA structure and function. The feature that underlies all aspects of inherited traits is the DNA sequence. For this reason, analyzing and altering DNA sequences is a powerful approach for understanding genetics. In this last section, we begin by examining a technique called **DNA sequencing.** This method enables researchers to determine the base sequence of DNA found in genes and other chromosomal regions. It is one of the most important tools for exploring genetics at the molecular level. DNA sequencing is practiced by scientists around the world, and the amount of scientific information contained within experimentally determined DNA sequences has become enormous. In Chapter 21, we will learn how researchers can determine the complete DNA sequence of entire genomes. In Chapter 22, we will consider how computers play an essential role in the storage and analysis of genetic sequences.

Not only can researchers determine DNA sequences, they may also use another technique, known as site-directed mutagenesis, to change the sequence of cloned DNA segments. At the end of this section, we will examine how site-directed mutagenesis is conducted and how it provides information regarding the function of genes.

The Dideoxy Method of DNA Sequencing Is Based on Our Knowledge of DNA Replication

Molecular geneticists often want to determine DNA base sequences as a first step toward understanding the function and expression of genes. For example, the investigation of genetic sequences has been vital to our understanding of promoters, regulatory elements, and the genetic code itself. Likewise, an examination of sequences has facilitated our understanding of origins of replication, centromeres, telomeres, and transposable elements.

During the 1970s, two methods for DNA sequencing were devised. One method, developed by Allan Maxam and Walter Gilbert, involved the base-specific chemical cleavage of DNA. Another method, developed by Frederick Sanger and colleagues, is known as **dideoxy sequencing.** Because it has become the more popular method of DNA sequencing, we consider the dideoxy method here. In addition, Chapter 21 considers some newer methods of sequencing DNA that are not based on the Sanger dideoxy method. These newer methods are commonly used in projects aimed at determining the DNA sequence of entire genomes.

The dideoxy procedure of DNA sequencing is based on our knowledge of DNA replication but uses a clever twist. As described in Chapter 13, DNA polymerase connects adjacent deoxyribonucleotides by catalyzing a covalent bond between the 5′ phosphate on one nucleotide and the 3′ –OH group on the previous nucleotide (refer back to Figure 13.15). Chemists, though, can synthesize deoxyribonucleotides that are missing the –OH group at the 3′ position (**Figure 19.16**). These synthetic nucleotides are called **dideoxyribonucleotides (ddNTPs).** (Note: The prefix dideoxy- indicates that two (di) oxygens (oxy) are removed (de) from this sugar compared with ribose; ribose has –OH groups at both the 2′ and 3′ positions.) Sanger reasoned that if a dideoxyribonucleotide is added to a growing DNA strand, the strand can no longer grow because the dideoxyribonucleotide is missing the 3′ –OH group. The incorporation of a dideoxyribonucleotide into a growing strand is therefore referred to as **chain termination.**

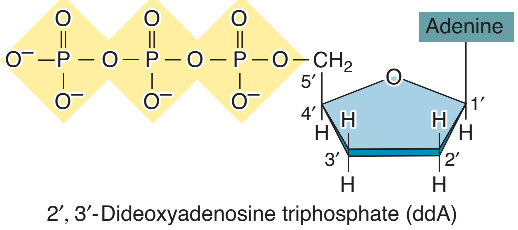


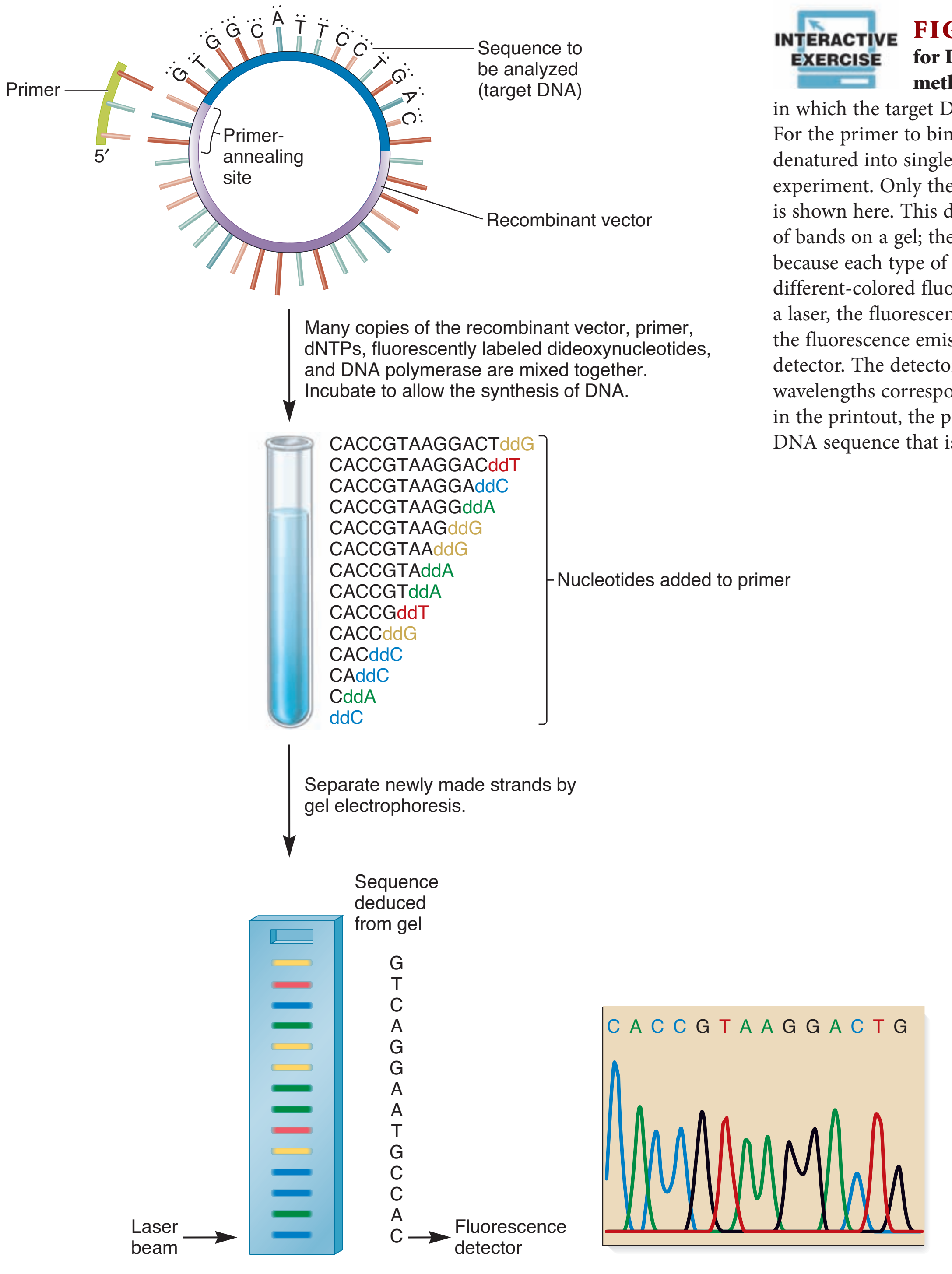
FIGURE 19.16 **The structure of a dideoxyribonucleotide.** Note that the 3′ group is a hydrogen rather than an –OH group. For this reason, another nucleotide cannot be attached at the 3′ position.

To detect the incorporation of dideoxynucleotides during DNA replication, the newly made DNA strands must be labeled in some way. When dideoxy sequencing was first invented, researchers labeled the newly made DNA with radioisotopes, which allowed them to be detected via autoradiography. This traditional DNA sequencing method has been largely replaced by **automated DNA sequencing** in which each type of dideoxynucleotide is labeled with a different colored fluorescent molecule. For example, a common way to fluorescently label ddNTPs is the following: ddA is green, ddT is red, ddG is yellow, and ddC is blue.

For automated DNA sequencing, the segment of DNA to be sequenced must be obtained in large amounts. This is accomplished using gene cloning, which was described earlier in this chapter. In the example in **Figure 19.17**, the segment of DNA to be sequenced, which we will call the target DNA, was cloned into a vector at a defined location. The target DNA was inserted next to a site in the vector where a primer will bind, which is called the primer-annealing site. The aim of the experiment is to determine the base sequence of the target DNA. In the experiment shown in Figure 19.17, the recombinant vector DNA has been previously denatured into single strands, usually via heat treatment. Only the strand needed for DNA sequencing is shown here.

Let's now consider the steps that are involved in DNA sequencing (see Figure 19.17). First, a sample containing many copies of the single-stranded DNA is mixed with many primers that will bind to the primer-annealing site. The primer binds to the DNA because the primer and primer-annealing site are complementary to each other. All four types of deoxyribonucleotides and DNA polymerase are added to the annealed DNA fragments. The tube also has a low concentration of each dideoxyribonucleotide (ddG, ddA, ddT, or ddC), which are fluorescently labeled. The tube is then incubated to allow DNA polymerase to make strands complementary to the target DNA sequence.

DNA synthesis continues until a dideoxynucleotide is incorporated into a growing strand. For example, chain termination can occasionally occur at the sixth or thirteenth position


G T G G C A T T C C T G A C
Sequence to be analyzed (target DNA)
Primer
5′
Primer-annealing site
Recombinant vector
Many copies of the recombinant vector, primer, dNTPs, fluorescently labeled dideoxynucleotides, and DNA polymerase are mixed together. Incubate to allow the synthesis of DNA.
CACCGTAAGGACTddG
CACCGTAAGGACddT
CACCGTAAGGAddC
CACCGTAAGGddA
CACCGTAAGddG
CACCGTAAddG
CACCGTAddA
CACCGTddA
CACCGddT
CACCddG
CACddC
CAddC
CddA
ddC
Nucleotides added to primer
Separate newly made strands by gel electrophoresis.
Sequence deduced from gel
G T C A G G A A T G C C A C
Laser beam
Fluorescence detector
C A C C G T A A G G A C T G

(a) Automated DNA sequencing

(b) Output from automated sequencing

INTERACTIVE EXERCISE **FIGURE 19.17 The protocol for DNA sequencing by the dideoxy method. (a)** The method begins with DNA in which the target DNA has been inserted into a vector. For the primer to bind, the recombinant vector must be denatured into single-stranded DNA at the beginning of the experiment. Only the strand needed for DNA sequencing is shown here. This diagram schematically depicts a series of bands on a gel; the four colors of the bands occur because each type of dideoxynucleotide is labeled with a different-colored fluorescent molecule. As each band passes a laser, the fluorescent dye is excited by the laser beam, and the fluorescence emission is recorded by a fluorescence detector. The detector reads the level of fluorescence at four wavelengths corresponding to the four dyes. **(b)** As shown in the printout, the peaks of fluorescence correspond to the DNA sequence that is complementary to the target DNA.

of the newly synthesized DNA strand if a ddT becomes incorporated at either of these sites. Note that the complementary A base is found at the sixth and thirteenth position in the target DNA. Therefore, we expect to make DNA strands that terminate at the sixth or thirteenth positions and have a ddT at their ends. Because these DNA strands contain a ddT, they are fluorescently labeled in red. Alternatively, ddA causes chain termination at the second, seventh, eighth, or eleventh positions because a complementary T base is found at the corresponding positions in the target strand. Strands that are terminated with ddA are fluorescently labeled in green.

After the samples have been incubated for several minutes, mixtures of DNA strands of different lengths are made, depending on the number of nucleotides attached to the primer. These DNA strands can be separated according to their lengths by running them on a slab gel or more commonly by running them through a gel-filled capillary tube. The shorter strands move to the bottom of the gel more quickly than the longer strands. Because we know the color of each dideoxyribonucleotide, we also know which base is at the very end of each DNA strand separated on the gel. Therefore, we can deduce the DNA sequence that is complementary to the target DNA by "reading" which base is at the end of every DNA strand and matching this sequence with the length of the strand. Reading the base sequence, from bottom to top, is much like climbing a ladder of bands. For this reason, the sequence obtained by this method is referred to as a **sequencing ladder.**

Theoretically, it is possible to read this sequence directly from the gel. From a practical perspective, however, it is faster and more efficient to automate the procedure using a laser and fluorescence detector. As the gel is running, each band passes the laser and the laser beam excites the fluorescent dye. The fluorescence detector records the amount of fluorescence emission from the excited dye. The detector reads the level of fluorescence at four wavelengths corresponding to the four different-colored dyes. An example of the printout from the fluorescence detector is shown in **Figure 19.17b**. As seen here, the peaks of fluorescence correspond to the DNA sequence that is complementary to the target DNA. Note that ddG is usually labeled with a yellow dye, but it is converted to black ink on the printout shown in Figure 19.17b for ease of reading. Though improvements in automated sequencing continue to be made, a typical sequencing run can provide a DNA sequence that is 700 to 900 bases long, and perhaps even longer.

Site-Directed Mutagenesis Is a Technique for Altering DNA Sequences

As we have seen, dideoxy sequencing provides a way of determining the base sequence of DNA. To understand how the genetic material functions, researchers often analyze mutations that alter the normal DNA sequence, thereby affecting the expression of genes and the outcome of traits. For example, geneticists have discovered that many inherited human diseases, such as sickle cell disease and hemophilia, involve mutations within specific genes. These mutations provide insight into the function of the genes in unaffected individuals. Hemophilia, for example, involves deleterious mutations in genes that encode blood-clotting factors.

Because the analysis of mutations can provide important information about normal genetic processes, researchers often wish to obtain mutant organisms. As we discussed in Chapter 18, mutations can arise spontaneously or can be induced by environmental agents. Mendel's pea plants are a classic example of allelic strains with different phenotypes that arose from spontaneous mutations. X-rays and UV light are physical agents that cause induced mutations. In addition, experimental organisms can be treated with chemical mutagens that increase the rate of mutation.

More recently, researchers have developed molecular techniques for making mutations within cloned genes or other DNA segments. One widely used method, known as **site-directed mutagenesis,** allows a researcher to produce a mutation at a specific site within a cloned DNA segment. For example, if a DNA sequence is 5′–AAATTTCTTTAAA–3′, a researcher can use site-directed mutagenesis to change it to 5′–AAATTTGTTTAAA–3′. In this case, the researcher deliberately changed the seventh base from a C to a G. Why is this method useful? The site-directed mutant can then be introduced into a living organism to see how the mutation affects the expression of a gene, the function of a protein, and the phenotype of an organism.

The first successful attempts at site-directed mutagenesis involved changes in the sequences of viral genomes. These studies were conducted in the 1970s. Mark Zoller and Michael Smith also developed a protocol for the site-directed mutagenesis of DNA that has been cloned into a viral vector. Since these early studies, many approaches have been used to achieve site-directed mutagenesis. **Figure 19.18** describes the general steps in the procedure. Prior to this experiment, the DNA was denatured into single strands; only the single strand needed for site-directed mutagenesis is shown. As in PCR, this single-stranded DNA is referred to as the template DNA, because it is used as a template to synthesize a complementary strand.

As shown in Figure 19.18, an oligonucleotide primer is allowed to hybridize or anneal to the template DNA. The primer, typically 20 or so nucleotides in length, is synthesized chemically. (A shorter version of the primer is shown in Figure 19.18 for simplicity.) The scientist designs the base sequence of the primer. The primer has two important characteristics. First, most of the sequence of the primer is complementary to the site in the DNA where the mutation is to be made. However, a second feature is that the primer contains a region of mismatch where the primer and template DNA are not complementary. The mutation occurs in this mismatched region. For this reason, site-directed mutagenesis is sometimes referred to as oligonucleotide-directed mutagenesis.

After the primer and template have annealed, the complementary strand is synthesized by adding deoxyribonucleoside triphosphates (dNTPs), DNA polymerase, and DNA ligase. This yields a double-stranded molecule that contains a mismatch only at the desired location. This double-stranded DNA can then be introduced into a bacterial cell. Within the cell, the DNA mismatch is likely to be repaired (see Chapter 18). Depending on which base is replaced (either C or A in the figure), this may produce a mutant sequence or the original sequence. Clones containing the desired mutation are identified by DNA sequencing and used for further studies.

After a site-directed mutation has been made within a cloned gene, its consequences are analyzed by introducing the mutant gene into a living cell or organism. As described earlier, recombinant vectors containing cloned genes can be introduced into bacterial cells. Following transformation into a bacterium, a researcher can study the differences in function between the mutant and wild-type genes and the proteins they encode. Similarly, mutant genes made via site-directed mutagenesis can be introduced into plants and animals, as discussed in Chapter 20.

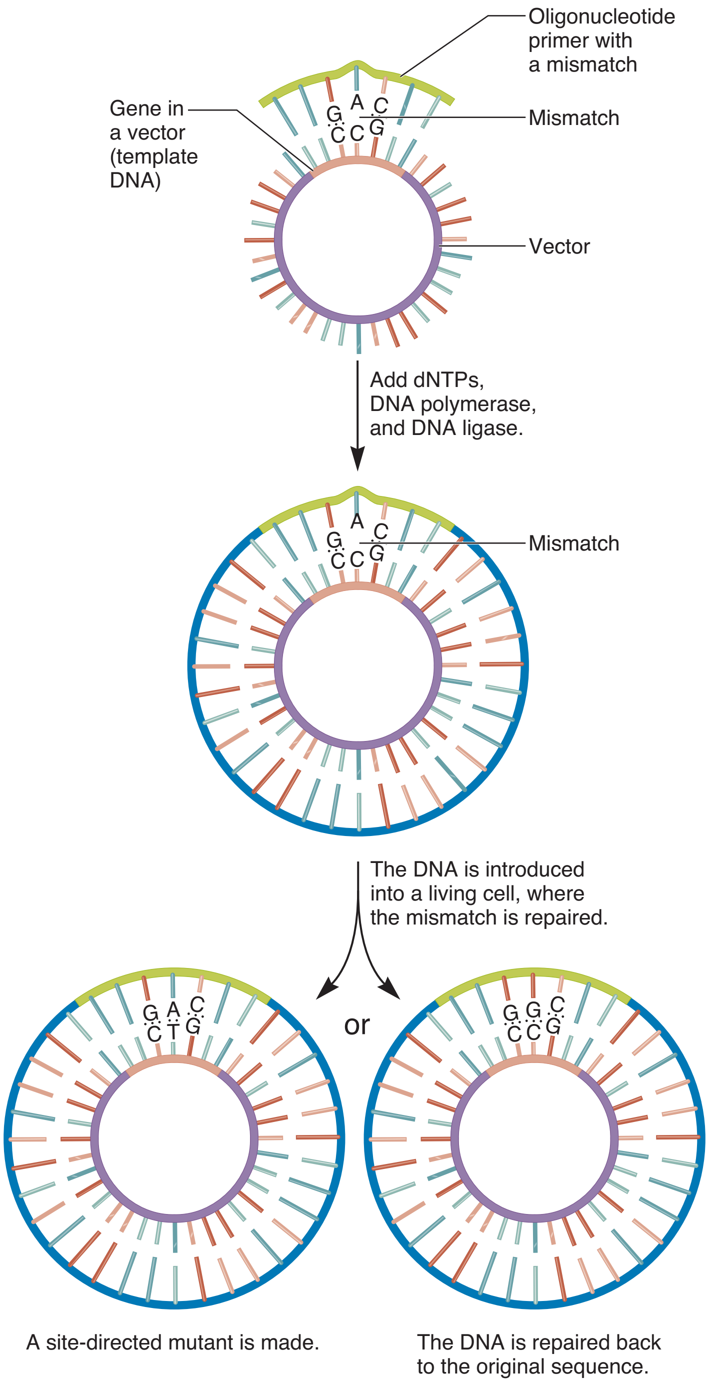

Oligonucleotide primer with a mismatch
Gene in a vector (template DNA)
G A C
C C G
Mismatch
Vector
Add dNTPs, DNA polymerase, and DNA ligase.
G A C
C C G
Mismatch
The DNA is introduced into a living cell, where the mismatch is repaired.
G A C
C T G
or
G G C
C C G
A site-directed mutant is made.
The DNA is repaired back to the original sequence.

FIGURE 19.18 **The method of site-directed mutagenesis.**

Genes→Traits To examine the relationship between genes and traits, researchers can alter gene sequences via site-directed mutagenesis. The altered gene can then be introduced into a living organism to examine how the mutation affects the organism's traits. For example, a researcher could introduce a nonsense mutation into the middle of the *lacY* gene in the *lac* operon. If this site-directed mutant was introduced into an *E. coli* bacterium that did not a have a normal copy of the *lacY* gene, the bacterium would be unable to use lactose. These results indicate that a functional *lacY* gene is necessary for bacteria to have the trait of lactose utilization.

Concept check: Describe three possible uses of site-directed mutagenesis.

19.5 REVIEWING THE KEY CONCEPTS

- The most commonly used method of DNA sequencing, called dideoxy sequencing, uses fluorescently labeled dideoxynucleotides that cause chain termination (see Figures 19.16, 19.17).
- In the method of site-directed mutagenesis, an oligonucleotide with one or more mismatches directs a mutation into a specific region of DNA (see Figure 19.18).

19.5 COMPREHENSION QUESTIONS

1. When a dideoxynucleotide is incorporated into a growing DNA strand,
   a. the strand elongates faster.
   b. the strand cannot elongate.
   c. the strand becomes more susceptible to DNase I cleavage.
   d. none of the above.
2. The purpose of site-directed mutagenesis is to
   a. determine if a protein binds to a DNA segment.
   b. determine the sequence of a segment of DNA.
   c. alter the sequence of a segment of DNA.
   d. determine if a gene is expressed.

KEY TERMS

Page 443. recombinant DNA technology, recombinant DNA molecules

Page 444. gene cloning, vector

Page 445. host cell, plasmids, R factors, origin of replication, selectable marker, shuttle vector, expression vector

Page 446. restriction endonucleases, restriction enzymes

Page 447. DNA ligase, palindromic, recombinant vector, competent cells, transformation

Page 449. reverse transcriptase, oligonucleotide

Page 450. complementary DNA (cDNA), polymerase chain reaction (PCR), primers

Page 451. template DNA, *Taq* polymerase, annealing, primer extension

Page 452. thermocycler

Page 453. reverse transcriptase PCR, real-time PCR

Page 455. cycle threshold method (C_t method), DNA library, genomic library

Page 457. cDNA library, colony hybridization, probe

Page 458. Southern blotting

Page 460. high stringency, low stringency, Northern blotting, Western blotting, antibody, epitopes, antigen

Page 462. gel retardation assay, gel mobility shift assay, DNase I footprinting

Page 464. DNA sequencing, dideoxy sequencing, dideoxyribonucleotides (ddNTPs), chain termination, automated DNA sequencing

Page 466. sequencing ladder, site-directed mutagenesis

CHAPTER SUMMARY

- Recombinant DNA technology is the use of in vitro molecular techniques to manipulate fragments of DNA to produce new arrangements.

19.1 Gene Cloning Using Vectors

- Gene cloning has many uses, including DNA sequencing, site-directed mutagenesis, the use of genes as probes, and the expression of cloned genes (see Table 19.1).
- Cloning vectors are derived from plasmids or viruses (see Table 19.2).
- Restriction enzymes, also called restriction endonucleases, cut chromosomal DNA and vector DNA to produce sticky ends that can hydrogen bond with each other. DNA ligase is needed to make a covalent link in the DNA backbone (see Figure 19.1, Table 19.3).
- Gene cloning using vectors involves the insertion of the gene into a vector and then its propagation in a living cell such as *E. coli* (see Figure 19.2).
- Complementary DNA (cDNA) can be made via reverse transcriptase by starting with mRNA (see Figure 19.3).

19.2 Polymerase Chain Reaction

- Polymerase chain reaction (PCR) uses oligonucleotide primers to copy a specific region of DNA. Each cycle of PCR involves three steps: denaturation, primer annealing, and primer extension (see Figures 19.4, 19.5).
- Reverse transcriptase PCR is used to detect and quantitate the amount of an RNA of interest in living cells (see Figure 19.6).
- Real-time PCR monitors PCR as it occurs in a thermocycler. It can be used to quantitatively measure the amount of starting DNA or mRNA in a sample. Real-time PCR uses fluorescent detectors to follow the PCR reaction. A sample that has an unknown amount of DNA is compared with a standard (see Figures 19.7, 19.8).

19.3 DNA Libraries and Blotting Methods

- A DNA library is a collection of recombinant vectors. When the starting material is chromosomal DNA, the library is called a genomic library; when the starting material is mRNA, it is called a cDNA library (see Figure 19.9).
- Colony hybridization is used to identify colonies that carry a gene or cDNA of interest (see Figure 19.10).
- Southern blotting uses a labeled DNA probe to detect the presence of a particular gene sequence within a mixture of many gene sequences. At low stringency, gene families may be detected (see Figure 19.11).
- Northern blotting uses a labeled DNA probe to detect a specific RNA within a mixture of many different RNAs (see Figure 19.12).
- Western blotting uses antibodies to detect a specific protein within a mixture of many different proteins (see Figure 19.13).

19.4 Methods for Analyzing DNA- and RNA-Binding Proteins

- A gel retardation assay can determine if a protein binds to a specific DNA fragment or RNA molecule because the binding of the protein slows down the movement of the DNA or RNA through a gel (see Figure 19.14).
- DNase I footprinting can determine the regions of a DNA molecule that are bound by a protein (see Figure 19.15).

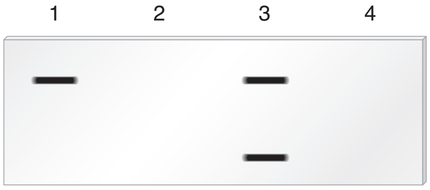
19.5 DNA Sequencing and Site-Directed Mutagenesis

- The most commonly used method of DNA sequencing, called dideoxy sequencing, uses fluorescently labeled dideoxynucleotides that cause chain termination (see Figures 19.16, 19.17).
- In the method of site-directed mutagenesis, an oligonucleotide with one or more mismatches directs a mutation into a specific region of DNA (see Figure 19.18).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. RNA was isolated from four different cell types and probed with radiolabeled DNA strands from a cloned gene that is called gene *X*. The results are shown here.

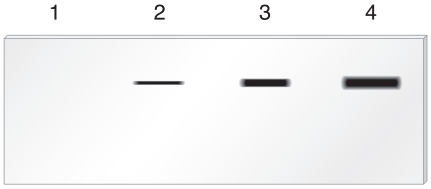


Lane 1: Muscle cells
Lane 2: Liver cells
Lane 3: Spleen cells
Lane 4: Nerve cells

Explain the results of this experiment.

Answer: In this Northern blot, a dark band appears in those lanes where RNA was isolated from muscle and spleen cells but not from liver and nerve cells. These results indicate that the muscle and spleen cells contain a significant amount of RNA from gene *X*, but the liver and nerve cells do not. The muscle cells show a single band, whereas the spleen cells show this band plus a second band of lower molecular mass. An interpretation of these results is that the spleen cells alternatively splice the RNA to produce a second RNA containing fewer exons.

S2. In the Western blotting experiment shown here (see next page), proteins were extracted from red blood cells obtained from tissue samples at different stages of human development. An equal amount of total cellular proteins was added to each lane. The primary antibody recognizes the β-globin polypeptide that is found in the hemoglobin protein.



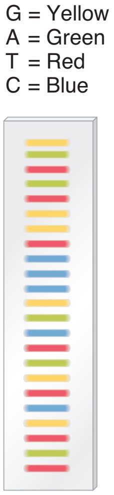
Lane 1: Embryonic red blood cells
Lane 2: Fetal red blood cells
Lane 3: Newborn red blood cells
Lane 4: Adult red blood cells

Explain these results.

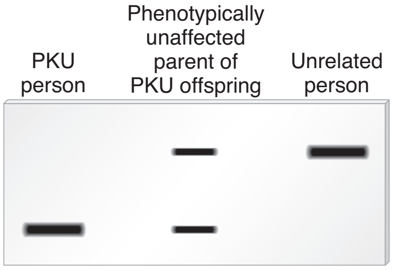
Answer: As shown here, the amount of β globin increases during development. Little detectable β globin is produced during embryonic development. The amount increases significantly during fetal development and becomes maximal in the adult. These results indicate that the β-globin gene is "turned on" in later stages of development, leading to the synthesis of the β-globin polypeptide. This experiment illustrates how a Western blot can provide information concerning the relative amount of a specific protein within living cells.

S3. A portion of a DNA strand is 3′–ATACGACTAGTCGGGACCA TATC–5′. If the primer in a dideoxy sequencing reaction anneals just to the left of this sequence, draw what the sequencing ladder would look like.

Answer:



S4. The human genetic disease phenylketonuria (PKU) involves a defect in a gene that encodes the enzyme phenylalanine hydroxylase. It is inherited as a recessive autosomal disorder. Using a strand of DNA from the nonmutant phenylalanine hydroxylase gene as a probe, a Southern blot was carried out on a PKU patient, one of her parents, and an unaffected, unrelated person. In this example, the DNA fragments were subjected to acrylamide gel electrophoresis, rather than agarose gel electrophoresis, because acrylamide gel electrophoresis is better able to detect small deletions within genes. The following results were obtained:



Suggest an explanation for these results.

Answer: In the affected person, the PKU defect is caused by a small deletion within the PKU gene. The unaffected parent is heterozygous for the nonmutant gene and the deletion. The PKU-affected person carries only the deletion, which runs at a lower molecular mass than the nonmutant gene. The unrelated person carries two copies of the nonmutant (nondeleted) gene.

Conceptual Questions

C1. Discuss three important advances that have resulted from gene cloning.

C2. What is a restriction enzyme? What structure does it recognize? What type of chemical bond does it cleave? Be as specific as possible.

C3. Write a double-stranded sequence that is 20 bases long and is palindromic.

C4. What is cDNA? In eukaryotes, how does cDNA differ from genomic DNA?

C5. Explain and draw the structural feature of a dideoxyribonucleotide that causes chain termination.

Application and Experimental Questions

E1. What is the functional significance of sticky ends in a cloning experiment? What type of bonding makes the ends sticky?

E2. Table 19.3 describes the cleavage sites of six different restriction enzymes. After these restriction enzymes have cleaved the DNA, five of them produce sticky ends that can hydrogen bond with complementary sticky ends, as shown in Figure 19.1. The efficiency of sticky ends binding together depends on the number of hydrogen bonds; more hydrogen bonds makes the ends "stickier" and more likely to stay attached. Rank these five restriction enzymes in Table 19.3 (from best to worst) with regard to the efficiency of their sticky ends binding to each other.

E3. Describe the important features of cloning vectors. Explain the purpose of selectable marker genes in cloning experiments.

E4. How does gene cloning produce many copies of a gene?

E5. In your own words, describe the series of steps necessary to clone a gene. Your answer should include the use of a probe to identify a bacterial colony that contains the cloned gene of interest.

E6. What is a recombinant vector? How is a recombinant vector constructed? Explain how X-Gal can be used in a method of identifying recombinant vectors that contain segments of chromosomal DNA.

E7. If a researcher began with a sample that contained three copies of double-stranded DNA, how many copies would be present after 27 cycles of PCR if PCR was 100% efficient?

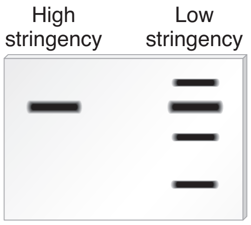
E8. Why is a thermostable form of DNA polymerase (e.g., *Taq* polymerase) used in PCR? Is it necessary to use a thermostable form of DNA polymerase in the techniques of dideoxy DNA sequencing or site-directed mutagenesis?

E9. Describe how you could produce many copies of β-globin cDNA from the rat using reverse transcriptase-PCR.

E10. What type of detector is used for real-time PCR? Explain how the level of fluorescence correlates with the level of PCR product.

E11. What phase of PCR (exponential, linear, or plateau) is analyzed to quantitate the amount of DNA or RNA in a sample? Explain why this phase is chosen.

E12. Let's suppose you have recently cloned a gene, which we will call gene *X*, from corn. You use a labeled DNA strand from this cloned gene to probe genomic DNA from corn in a Southern blot experiment under conditions of low and high stringency. The following results were obtained:



What do these results mean?

E13. What is a DNA library? Do you think this is an appropriate name?

E14. Some vectors used in cloning experiments contain bacterial promoters that are adjacent to unique cloning sites. This makes it possible to insert a gene sequence next to the bacterial promoter and express the gene in bacterial cells. These are called expression vectors. If you wanted to express a eukaryotic protein in bacterial cells, would you clone genomic DNA or cDNA into the expression vector? Explain your choice.

E15. Southern and Northern blotting depend on the phenomenon of hybridization. In these two techniques, explain why hybridization occurs. Which member of the hybrid is labeled?

E16. In Southern, Northern, and Western blotting, what is the purpose of gel electrophoresis?

E17. What is the purpose of a Northern blotting experiment? What types of information can it tell you about the transcription of a gene?

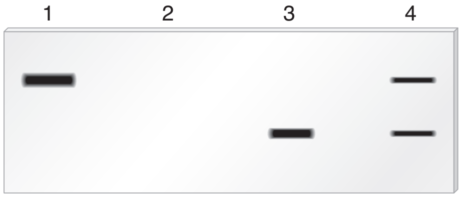
E18. Let's suppose an X-linked gene in mice exists as two alleles, which we will call *B* and *b*. X inactivation, a process in which one X chromosome is turned off, occurs in the somatic cells of female mammals (see Chapter 4). Allele *B* encodes an mRNA that is 900 nucleotides long, whereas allele *b* contains a small deletion that shortens the mRNA to a length of 825 nucleotides. Draw the expected results of a Northern blot using mRNA isolated from somatic tissue of the following mice:

Lane 1. mRNA from an X^bY male mouse

Lane 2. mRNA from an X^bX^b female mouse

Lane 3. mRNA from an X^BX^b female mouse. Note: The sample taken from the female mouse is not from a clone of cells.

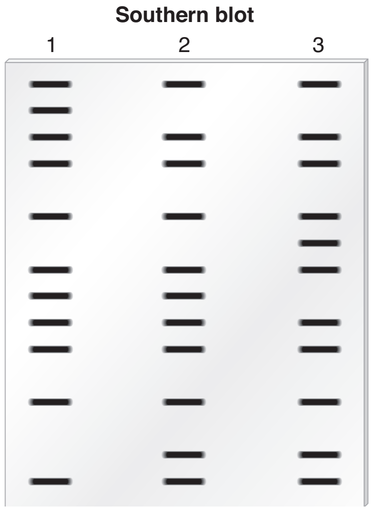
E19. The method of Northern blotting can be used to determine the amount and size of a particular RNA transcribed in a given cell type. Alternative splicing (discussed in Chapter 17) can produce mRNAs from the same gene that have different lengths. A Northern blot is shown here using a DNA probe that is complementary to the mRNA encoded by a particular gene. The mRNA in lanes 1 through 4 was isolated from different cell types, and equal amounts of total cellular mRNA were added to each lane.



Lane 1: mRNA isolated from nerve cells
Lane 2: mRNA isolated from kidney cells
Lane 3: mRNA isolated from spleen cells
Lane 4: mRNA isolated from muscle cells

Explain these results.

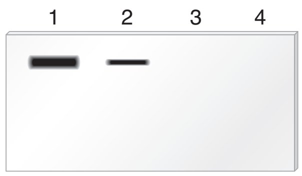
E20. Southern blotting can be used to detect the presence of repetitive sequences, such as transposable elements, that are present in multiple copies within the chromosomal DNA of an organism. (Note: Transposable elements are described in Chapter 21.) In the Southern blot shown here, chromosomal DNA was isolated from three different strains of baker's yeast, digested with a restriction enzyme, run on a gel, blotted, and then probed with a radioactive DNA probe that is complementary to a transposable element called the *Ty* element.



Explain, in a general way, why the banding patterns are not the same in lanes 1, 2, and 3.

E21. In Chapter 8, Figure 8.7 describes the evolution of the globin gene family. All of the genes in this family are homologous to each other, though the degree of sequence similarity varies depending on the time of divergence. Genes that have diverged more recently have sequences that are more similar. For example, the α_1 and α_2 genes have DNA sequences that are more similar to each other compared with the α_1 and ξ genes. In a Southern blotting experiment, the degree of sequence similarity can be discerned by varying the stringency of hybridization. At high temperature (i.e., high stringency), the probe recognizes genes that are only a perfect or very close match. At a lower temperature, however, homologous genes with lower degrees of similarities can be detected because slight mismatches are tolerated. If a Southern blot was conducted on a sample of human chromosomal DNA and a probe was used that was a perfect match to the β-globin gene, rank the following genes (from those that are detected at high stringency down to those that are only detected at low stringency) as they would appear in a Southern blot experiment: *Mb*, α_1, β, γ_A, δ, and ε.

E22. In the Western blot shown here, polypeptides were isolated from red blood cells and muscle cells from two different individuals. One individual was unaffected, and the other individual suffered from a disease known as thalassemia, which involves a defect in hemoglobin. In the Western blot, the gel blot was exposed to an antibody that recognizes β globin, which is one of the polypeptides that constitute hemoglobin. Equal amounts of total cellular proteins were added to each lane.



Lane 1: Proteins isolated from normal red blood cells
Lane 2: Proteins isolated from the red blood cells of a thalassemia patient
Lane 3: Proteins isolated from normal muscle cells
Lane 4: Proteins isolated from the muscle cells of a thalassemia patient

Explain these results.

E23. Let's suppose a researcher was interested in the effects of mutations on the expression of a structural gene that encodes a polypeptide 472 amino acids in length. This polypeptide is expressed in leaf cells of *Arabidopsis thaliana.* Because the average molecular mass of an amino acid is 120 Da, this protein has a molecular mass of approximately 56,640 Da. Make a drawing that shows the expected results of a Western blot using polypeptides isolated from the leaf cells that were obtained from the following individuals:

Lane 1. A plant homozygous for a nonmutant gene

Lane 2. A plant homozygous for a deletion that removes the promoter for this gene

Lane 3. A heterozygous plant in which one gene is nonmutant and the other gene has a mutation that introduces an early stop codon at codon 112

Lane 4. A plant homozygous for a mutation that introduces an early stop codon at codon 112

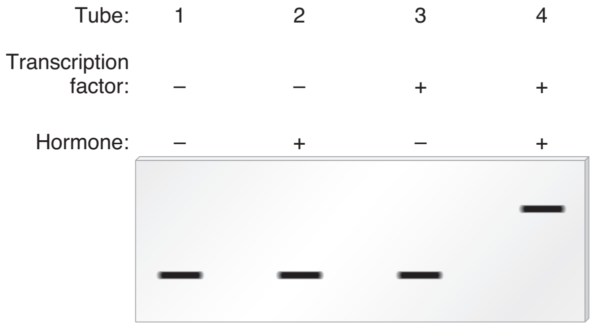
Lane 5. A plant homozygous for a mutation that changes codon 108 from a phenylalanine codon into a leucine codon

E24. If you wanted to know if a protein was made during a particular stage of development, what technique would you choose?

E25. Explain the basis for using an antibody as a probe in a Western blotting experiment.

E26. Starting with pig cells and a probe that is a labeled DNA strand from the human β-globin gene, describe how you would clone the β-globin gene from pigs. You may assume that you have available all of the materials needed in a cloning experiment. How would you confirm that a putative clone really contained the β-globin gene?

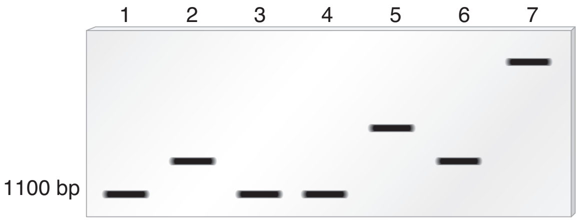
E27. A cloned gene fragment contains a response element that is recognized by a regulatory transcription factor. Previous experiments have shown that the presence of a hormone results in transcriptional activation by this transcription factor. To study this effect, you conduct a gel retardation assay and obtain the following results:



Explain the action of the hormone.

E28. Describe the rationale behind a gel retardation assay.

E29. A gel retardation assay can be used to study the binding of proteins to a segment of DNA. In the experiment shown here, a gel retardation assay was used to examine the requirements for the binding of RNA polymerase II (from eukaryotic cells) to the promoter of a structural gene. The assembly of general transcription factors and RNA polymerase II at the core promoter is described in Chapter 14 (Figure 14.14). In this experiment, the segment of DNA containing a promoter sequence was 1100 bp in length. The fragment was mixed with various combinations of proteins and then subjected to a gel retardation assay.



Lane 1: No proteins added
Lane 2: TFIID
Lane 3: TFIIB
Lane 4: RNA polymerase II
Lane 5: TFIID + TFIIB
Lane 6: TFIID + RNA polymerase II
Lane 7: TFIID + TFIIB + RNA polymerase II

Explain which proteins (TFIID, TFIIB, or RNA polymerase II) are able to bind to this DNA fragment by themselves. Which transcription factors (i.e., TFIID or TFIIB) are needed for the binding of RNA polymerase II?

E30. As described in Chapter 17 (Figure 17.7), certain regulatory transcription factors bind to the DNA and activate RNA polymerase II. When glucocorticoid binds to the glucocorticoid receptor (a regulatory transcription factor), this changes the conformation of the receptor and allows it to bind to the DNA. The glucocorticoid receptor binds to a DNA sequence called a glucocorticoid response element (GRE). In contrast, other regulatory transcription factors, such as the CREB protein, do not require hormone binding in order to bind to DNA. The CREB protein can bind to the DNA in the absence of any hormone, but it does not activate RNA polymerase II unless the CREB protein is phosphorylated. (Phosphorylation is stimulated by certain hormones.) The CREB protein binds to a DNA sequence called a cAMP response element (CRE). With these ideas in mind, draw the expected results of a gel retardation assay conducted on the following samples:

Lane 1. A 600-bp fragment containing a GRE, plus the glucocorticoid receptor

Lane 2. A 600-bp fragment containing a GRE, plus the glucocorticoid receptor, plus glucocorticoid hormone

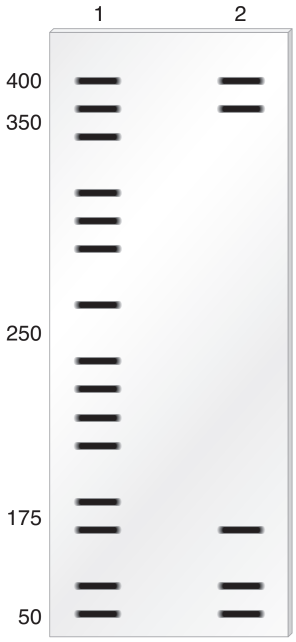
Lane 3. A 600-bp fragment containing a GRE, plus the CREB protein

Lane 4. A 700-bp fragment containing a CRE, plus the CREB protein

Lane 5. A 700-bp fragment containing a CRE, plus the CREB protein, plus a hormone (e.g., epinephrine) that causes the phosphorylation of the CREB protein

Lane 6. A 700-bp fragment containing a CRE, plus the glucocorticoid receptor, plus glucocorticoid hormone.

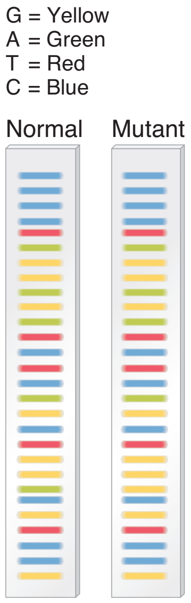
E31. In the technique of DNase I footprinting, the binding of a protein to a region of DNA protects that region from digestion by DNase I by blocking the ability of DNase I to gain access to the phosphodiester linkages in the DNA. In the DNase I footprinting experiment shown here, a researcher began with a sample of cloned DNA 400 bp in length. This DNA contained a eukaryotic promoter for RNA polymerase II. The assembly of general transcription factors and RNA polymerase II at the core promoter is described in Chapter 14 (Figure 14.14). For the sample loaded in lane 1, no proteins were added. For the sample loaded in lane 2, the 400-bp fragment was mixed with RNA polymerase II plus TFIID and TFIIB.



Which region of this 400-bp fragment of DNA is bound by RNA polymerase II and TFIID and TFIIB?

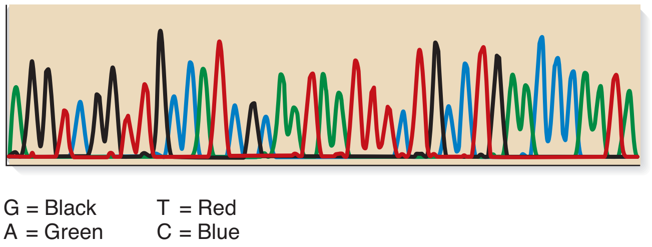
E32. Explain the rationale behind a DNase I footprinting experiment.

E33. DNA sequencing can help us to identify mutations within genes. The following data are derived from an experiment in which a normal gene and a mutant gene have been sequenced:


G = Yellow
A = Green
T = Red
C = Blue

Locate and describe the mutation.

E34. A sample of DNA was subjected to automated DNA sequencing as shown here.



What is the sequence of this DNA segment?

E35. A portion of the coding sequence of a cloned gene is shown here:

5′–GCCCCCGATCTACATCATTACGGCGAT–3′
3′–CGGGGGCTAGATGTAGTAATGCCGCTA–5′

This portion of the gene encodes a polypeptide with the amino acid sequence alanine–proline–aspartic acid–leucine–histidine–histidine–tyrosine–glycine–aspartic acid. Using the method of site-directed mutagenesis, a researcher wants to change the leucine codon into an arginine codon, with an oligonucleotide that is 19 nucleotides long. What is the sequence of the oligonucleotide that should be made? You should designate the 5′ and 3′ ends of the oligonucleotide in your answer. Note: The mismatch should be in the middle of the oligonucleotide, and a one-base mismatch is preferable over a two- or three-base mismatch. Use the bottom strand as the template strand for this site-directed mutagenesis experiment.

E36. Let's suppose you want to use site-directed mutagenesis to investigate a DNA sequence that functions as a response element for hormone binding. From previous work, you have narrowed down the response element to a sequence of DNA that is 20 bp in length with the following sequence:

5′–GGACTGACTTATCCATCGGT–3′
3′–CCTGACTGAATAGGTAGCCA–5′

As a strategy to pinpoint the actual response element sequence, you decide to make 10 different site-directed mutants and then analyze their effects by a gel retardation assay. What mutations would you make? What results would you expect to obtain?

E37. Site-directed mutagenesis is used to explore the structure and function of proteins. For example, changes can be made to the coding sequence of a gene to determine how alterations in the amino acid sequence affect the function of a protein. Let's suppose that you are interested in the functional importance of one glutamic acid residue within a protein you are studying. By site-directed mutagenesis, you make mutant proteins in which this glutamic acid codon has been changed to other codons. You then test the encoded mutant proteins for functionality. The results are as follows:

	Functionality
Normal protein	100%
Mutant proteins containing	
Tyrosine	5%
Phenylalanine	3%
Aspartic acid	94%
Glycine	4%

From these results, what would you conclude about the functional significance of the glutamic acid residue within the protein?

Questions for Student Discussion/Collaboration

1. Discuss and make a list of some of the reasons why it would be informative for a geneticist to determine the amount of a gene product. Use specific examples of known genes (e.g., β globin and other genes) when making your list.
2. Make a list of all the possible genetic questions that could be answered using site-directed mutagenesis.

Answers to Comprehension Questions

19.1: d, b, c, a
19.2: c, b, d
19.3: a, d, b, b
19.4: b, c
19.5: b, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

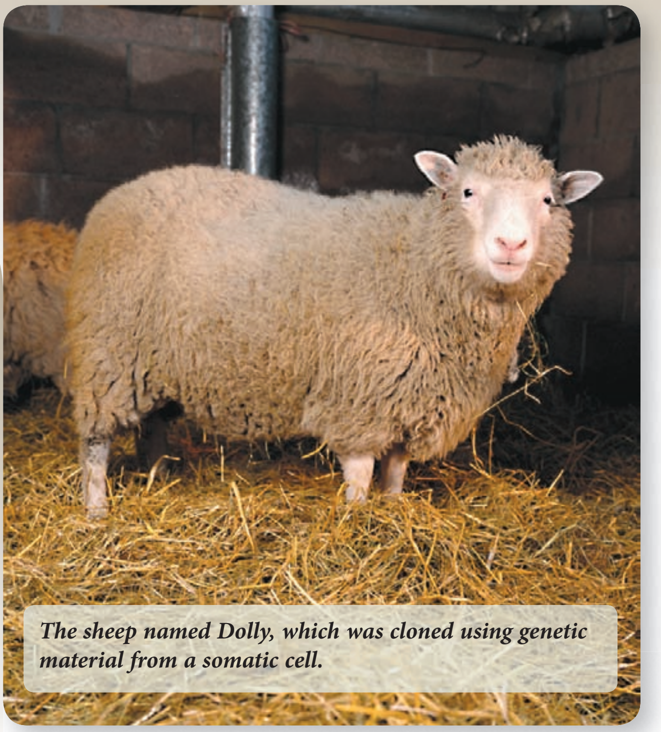
www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

20

CHAPTER OUTLINE

20.1 Uses of Microorganisms in Biotechnology
20.2 Genetically Modified Animals
20.3 Reproductive Cloning and Stem Cells
20.4 Genetically Modified Plants
20.5 Human Gene Therapy



The sheep named Dolly, which was cloned using genetic material from a somatic cell.

BIOTECHNOLOGY

Biotechnology is broadly defined as the application of technologies that involve the use of living organisms, or products from living organisms, for the development of products that benefit humans. Biotechnology is not a new topic. It began several thousand years ago when humans began to domesticate animals and plants for the production of food. Since that time, many species of microorganisms, animals, and plants have become routinely used by people. More recently, the term *biotechnology* has become associated with molecular genetics. Since the 1970s, molecular genetic tools have provided novel ways of making use of living organisms for their products and services. As discussed in Chapter 19, recombinant DNA techniques can be used to genetically engineer microorganisms. In addition, recombinant methods enable the introduction of genetic material into animals and plants. **Genetically modified organisms (GMOs)** have received genetic material via recombinant DNA technology. If an organism has received genetic material from a different species, it is called a **transgenic organism.** A gene from one species that is introduced into another species is called a **transgene.**

In the 1980s, court rulings made it possible to patent recombinant organisms such as transgenic animals and plants. This was one factor that contributed to the growth of many biotechnology industries. In this chapter, we will examine how molecular techniques have expanded our knowledge of the genetic characteristics of commercially important species. We will also discuss examples in which recombinant microorganisms and transgenic animals and plants have been given characteristics that are useful in the treatment of disease or in agricultural production. These include recombinant bacteria that make human insulin, transgenic livestock that produce human proteins in their milk, and transgenic corn with insect resistance. In addition, the topics of mammalian cloning and stem cell research are examined from a technical point of view. Likewise, the current and potential use of human gene therapy—the introduction of cloned genes into living cells in the treatment of a disease—will be addressed. In the process, we will also touch upon some of the ethical issues associated with these technologies.

20.1 USES OF MICROORGANISMS IN BIOTECHNOLOGY

Microorganisms are used to benefit humans in various ways (**Table 20.1**). In this section, we will examine how molecular

TABLE 20.1

Common Uses of Microorganisms

Application	Examples
Production of medicines	Antibiotics
	Synthesis of human insulin in recombinant *E. coli*
Food fermentation	Cheese, yogurt, vinegar, wine, and beer
Biological control	Control of plant diseases, insect pests, and weeds
	Symbiotic nitrogen fixation
	Prevention of frost formation
Bioremediation	Cleanup of environmental pollutants such as petroleum hydrocarbons and synthetics that are difficult to degrade

genetic tools have become increasingly important for improving our use of microorganisms. Such tools can produce recombinant microorganisms with genes that have been manipulated in vitro. Why are recombinant organisms useful? Recombinant techniques can improve strains of microorganisms and have even yielded strains that make products not normally produced by microorganisms. For example, human genes have been introduced into bacteria to produce medically important products such as insulin and human growth hormone. As discussed in this section, several recombinant strains are in widespread use. However, in some areas of biotechnology and in some parts of the world, the commercialization of recombinant strains has proceeded very slowly. This is particularly true for applications in which recombinant microorganisms may be used to produce food products or where they are released into the environment. In such cases, safety and environmental concerns along with negative public perceptions have slowed or even halted the commercial use of recombinant microorganisms. Nevertheless, molecular genetic research continues, and many biotechnologists expect an expanding use of recombinant microbes in the future.

Many Important Medicines Are Produced by Recombinant Microorganisms

During the 1970s, geneticists became aware of the great potential of recombinant DNA technology to produce therapeutic agents for treating certain human diseases. Healthy individuals possess many different genes that encode short peptide and longer polypeptide hormones. Diseases can result when an individual is unable to produce these hormones.

In 1976, Robert Swanson and Herbert Boyer formed Genentech, Inc. The aspiration of this company was to engineer bacteria to synthesize useful products, particularly peptide and polypeptide hormones. Their first contract was with researchers Keiichi Itakura and Arthur Riggs. They were able to engineer a bacterial strain that produced somatostatin, a human hormone that inhibits the secretion of a number of other hormones, including growth hormone, insulin, and glucagon. Somatostatin was not chosen for its commercial potential. Rather, it was chosen because the researchers thought it would be technically less difficult to produce than other hormones. Somatostatin is very small (only 14 amino acids long), requires a short coding sequence, and it can be detected easily. Since this pioneering work, recombinant DNA technology has been used to develop bacterial strains that synthesize several other medical agents, a few of which are described in **Table 20.2**.

In 1982, the U.S. Food and Drug Administration approved the sale of the first genetically engineered drug, human insulin, which was produced by Genentech and marketed by Eli Lilly. Insulin, which is produced by the β cells of the pancreas, functions to regulate several physiological processes, including the uptake of glucose into fat and muscle cells. Persons with insulin-dependent diabetes cannot synthesize an adequate amount of insulin due to a loss of their β cells. Prior to 1982, insulin was isolated from pancreases removed from cattle and pigs. Unfortunately, in some cases, diabetic individuals became allergic to such insulin and had to use expensive combinations of insulin from other animals and human cadavers. Today, people with diabetes can use genetically engineered human insulin to treat the disease.

Insulin is a hormone composed of two polypeptide chains, called the A and B chains. To make this hormone using bacteria, the coding sequences of the A and B chains are inserted next to the coding sequence of a native *E. coli* protein, β-galactosidase (**Figure 20.1**). This creates a fusion protein comprising β-galactosidase and the A or B chain. This step is necessary because the A and B chains are rapidly degraded when expressed in bacterial cells by themselves. The fusion proteins, however, are not. How are the two fusion proteins used to make human insulin? After the fusion proteins are expressed in bacteria, they are isolated and then treated with cyanogen bromide (CNBr), which cleaves after a methionine (Met) that is found at the junction between β-galactosidase and the A or B chain. This cleavage step separates β-galactosidase from the A or B chain. The A and B chains are then purified and mixed together under conditions in which they refold and form disulfide bonds with each other to make an active insulin hormone, which has the same structure as the human-made hormone.

TABLE 20.2

Examples of Medical Agents Produced by Recombinant Microorganisms

Drug	Action	Treatment
Insulin	A hormone that promotes glucose uptake	For diabetes
Tissue plasminogen activator (TPA)	Dissolves blood clots	For heart attacks and other arterial occlusions
Superoxide dismutase	Antioxidant	Used following heart attacks to minimize tissue damage
Factor VIII	Blood-clotting factor	For a certain type of hemophilia
Renin inhibitor	Lowers blood pressure	For hypertension
Erythropoietin	Stimulates the production of red blood cells	For anemia

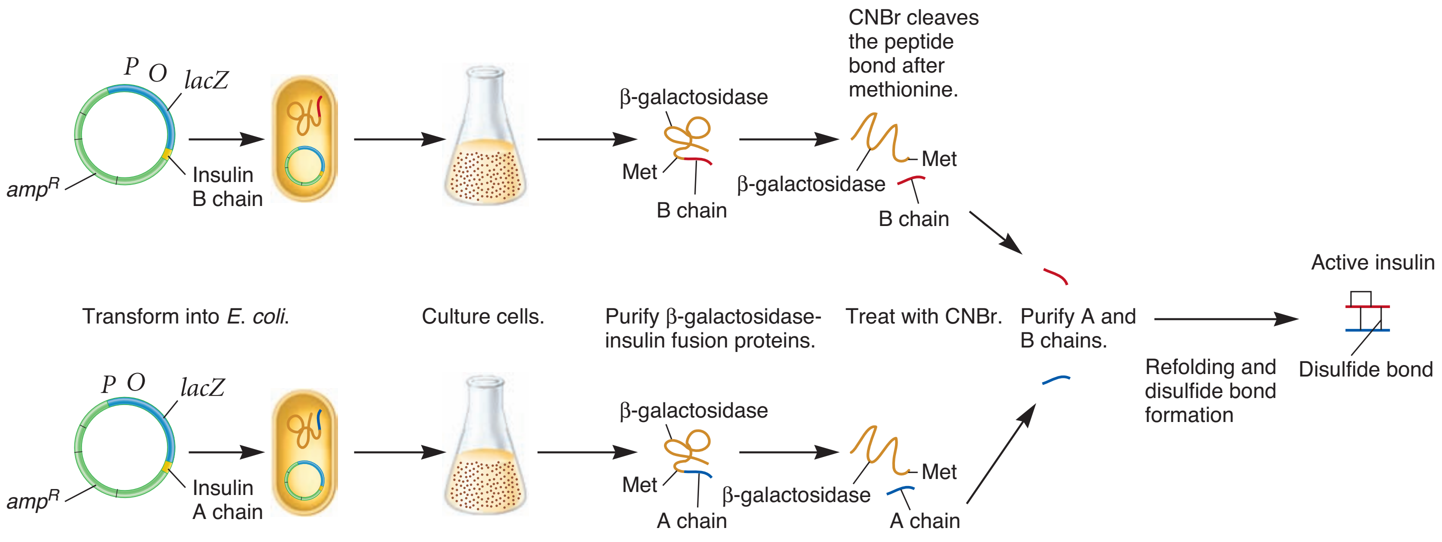


FIGURE 20.1 **The use of bacteria to make human insulin.**

Genes → Traits The synthesis of human insulin is not a trait that bacteria normally possess. However, genetic engineers introduced the genetic sequences that encode the A and B chains of human insulin via recombinant DNA technology, yielding bacteria that make these polypeptides as fusion proteins with β-galactosidase. Cyanogen bromide (CNBr) treatment releases the A and B polypeptides, which are then purified and mixed together under conditions in which they refold and form functional human insulin.

Concept check: *What is the purpose of using CNBr in this experiment?*

Bacterial Species Can Be Used as Biological Control Agents

The term **biological control** refers to the use of living organisms or their products to alleviate plant diseases and pests. During the past 20 years, interest in the biological control of plant diseases and insect pests as an alternative to chemical pesticides has increased. Biological control agents can prevent disease in several ways. In some cases, nonpathogenic microorganisms are used to compete effectively against pathogenic strains for nutrients or space. In other cases, microorganisms may produce a toxin that inhibits other pathogenic microorganisms or insects without harming the plant.

Biological control can also involve the use of microorganisms living in the field. A successful example is the use of *Agrobacterium radiobacter* to prevent crown gall disease caused by *Agrobacterium tumefaciens*. The disease gets its name from the large swellings (galls) produced by the plant in response to the bacteria. *A. radiobacter* produces agrocin 84, an antibiotic that kills *A. tumefaciens*. Molecular geneticists have determined that *A. radiobacter* contains a plasmid with genes that encode enzymes responsible for agrocin 84 synthesis, which is a small molecule. The plasmid also carries genes that confer resistance to this antibiotic. Unfortunately, this plasmid is occasionally transferred from *A. radiobacter* to *A. tumefaciens* during interspecies conjugation. When this occurs, *A. tumefaciens* gains resistance to agrocin 84. Researchers have identified *A. radiobacter* strains in which this plasmid has been altered genetically to prevent its transfer during conjugation. This conjugation-deficient strain is now used commercially worldwide to prevent crown gall disease.

Another biological control agent is *Bacillus thuringiensis*, usually referred to as Bt (pronounced “bee-tee”). This naturally occurring bacterium produces toxins that are lethal to many caterpillars and beetles that feed on a wide variety of food crops and ornamental plants. Bt is generally harmless to plants and other animals, such as humans, and does not usually harm beneficial insects that act as pollinators. Therefore, it is viewed as an environmentally friendly pesticide. Commercially, Bt is sold in a powder form that is used as a dust or mixed with water as a foliage spray. Bt is then dusted or sprayed on plants that are under attack by caterpillars or beetles so the pests will ingest the bacteria as they eat the leaves, flowers, or fruits. The toxins produced by Bt bring about paralysis of the insect’s digestive tract, causing it to stop feeding within hours and die within a few days. Geneticists have cloned the genes that encode Bt toxins, which are proteins. As discussed later, such genes have been introduced into crops, such as corn, to produce transgenic plants resistant to insect attack.

Microorganisms Can Reduce Environmental Pollutants

The term **bioremediation** refers to the use of living organisms or their products to decrease pollutants in the environment. As its name suggests, this is a biological remedy for pollution. During bioremediation via microorganisms, enzymes produced by a microorganism modify a toxic pollutant by altering or transforming its structure. This event is called **biotransformation.** In many cases, biotransformation results in **biodegradation,** in which the toxic pollutant is degraded, yielding less complex, nontoxic metabolites. Alternatively, biotransformation without biodegradation can also occur. For example, toxic heavy metals can often be rendered less toxic by oxidation or reduction reactions carried out by microorganisms. Another way to alter the toxicity of organic pollutants is by promoting polymerization, in which toxic monomers are covalently joined to form a chain. In many cases, polymerized toxic compounds are less likely to leach from the soil and, therefore, are less environmentally toxic than their parent compounds.

Since the early 1900s, microorganisms have been used in the treatment and degradation of sewage. More recently, the field of bioremediation has expanded into the treatment of hazardous and refractory chemical wastes—chemicals that are difficult to degrade and usually associated with industrial activity. These pollutants include petroleum hydrocarbons, halogenated organic compounds, pesticides, herbicides, and organic solvents. Many new applications that use microorganisms to degrade these pollutants are being tested. The field of bioremediation has been fostered, to a large extent, by better knowledge of how pollutants are degraded by microorganisms, the identification of new and useful strains of microbes, and the ability to enhance bioremediation through genetic engineering.

Molecular genetic technology is key in identifying genes that encode enzymes involved in bioremediation. The characterization of the relevant genes greatly enhances our understanding of how microbes can modify toxic pollutants. In addition, recombinant strains produced in the laboratory can be more efficient at degrading certain types of pollutants.

In 1980, in a landmark case (*Diamond v. Chakrabarty*), the U.S. Supreme Court ruled that a live, recombinant microorganism is patentable as a "manufacture or composition of matter." The first recombinant microorganism to be patented was an "oil-eating" bacterium that contained a laboratory-constructed plasmid. This strain can oxidize the hydrocarbons commonly found in petroleum. It grew faster on crude oil than did any of the natural isolates tested. However, it has not been a commercial success because this recombinant strain metabolizes only a limited number of toxic compounds, a fraction of the more than 3000 actually present in crude oil. Unfortunately, the recombinant strain did not degrade many higher-molecular-weight compounds, which tend to persist in the environment.

Thus far, most bioremediation has involved the use of natural microorganisms rather than recombinant ones. Currently, bioremediation should be considered a developing industry. Many studies are currently underway aimed at elucidating the mechanisms whereby microorganisms degrade toxic pollutants. In the future, recombinant microorganisms may provide an effective way to decrease the levels of toxic chemicals within our environment. However, this approach will require careful studies to demonstrate that recombinant organisms are effective at reducing pollutants, and are safe and able to survive when released into the environment.

20.1 REVIEWING THE KEY CONCEPTS

- Microorganisms have several uses in biotechnology (see Table 20.1).
- Recombinant microorganisms are used to make human hormones such as insulin (see Table 20.2, Figure 20.1).
- Microorganisms may be used as biological control agents. In bioremediation, living organisms are used to decrease pollutants in the environment.

20.1 COMPREHENSION QUESTIONS

1. Which of the following is a use of a microorganism in biotechnology?
   a. Production of medicines
   b. Food fermentation
   c. Biological control
   d. All of the above
2. What is the key reason why the A and B chains of insulin are made as fusion proteins with β-galactosidase?
   a. To make purification easier
   b. To prevent degradation of the A and B chains
   c. To be secreted from the cell
   d. All of the above
3. Which of the following was the first living organism to be patented?
   a. A strain of *E. coli* that makes somatostatin
   b. A strain of *E. coli* that makes insulin
   c. An oil-eating bacterium
   d. A strain of *B. thuringiensis* that makes an insecticide

20.2 GENETICALLY MODIFIED ANIMALS

As mentioned at the beginning of this chapter, transgenic organisms contain recombinant DNA from another species that has been integrated into their genome. A dramatic example of this is shown in **Figure 20.2**. In this case, the gene that encodes human growth hormone was introduced into the genome of a

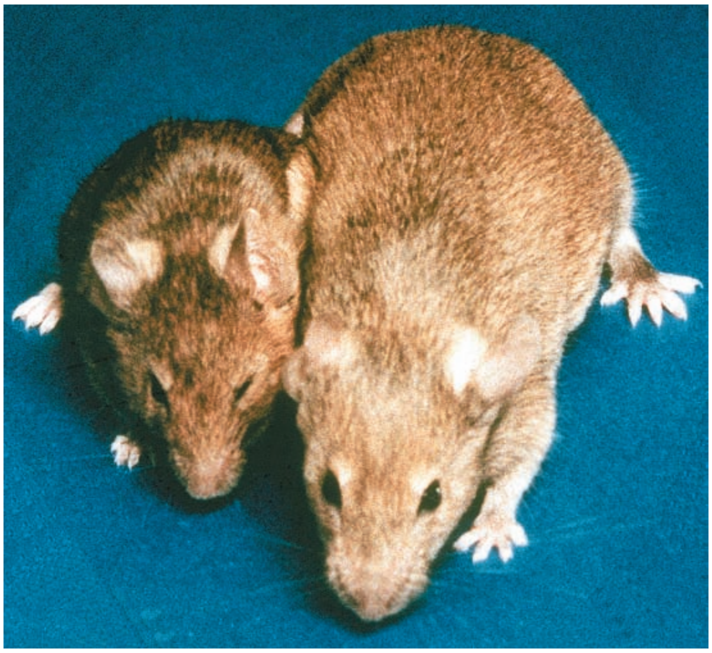


FIGURE 20.2 A comparison between a normal mouse and a transgenic mouse that carries a growth hormone gene.

Genes → Traits The transgenic mouse (on the right) carries a gene encoding human growth hormone. The introduction of the human gene into the mouse's genome has caused it to grow larger.

mouse. The larger mouse on the right is a transgenic mouse that expresses the human growth hormone gene to a very high level.

The production of transgenic animals is a relatively new area of biotechnology. In recent years, a few transgenic species have reached the stage of commercialization. Many researchers believe that this technology holds great promise for innovations in biotechnology. However, the degree to which this potential may be realized depends, in part, on the public's concern about the production and consumption of transgenic species.

In this section, we begin by examining the mechanisms through which cloned DNA becomes integrated into the chromosomal DNA of animal cells. We will then explore the current techniques used to produce transgenic animals and their potential uses.

The Integration of a Cloned Gene into a Chromosome Can Result in Gene Replacement or Gene Addition

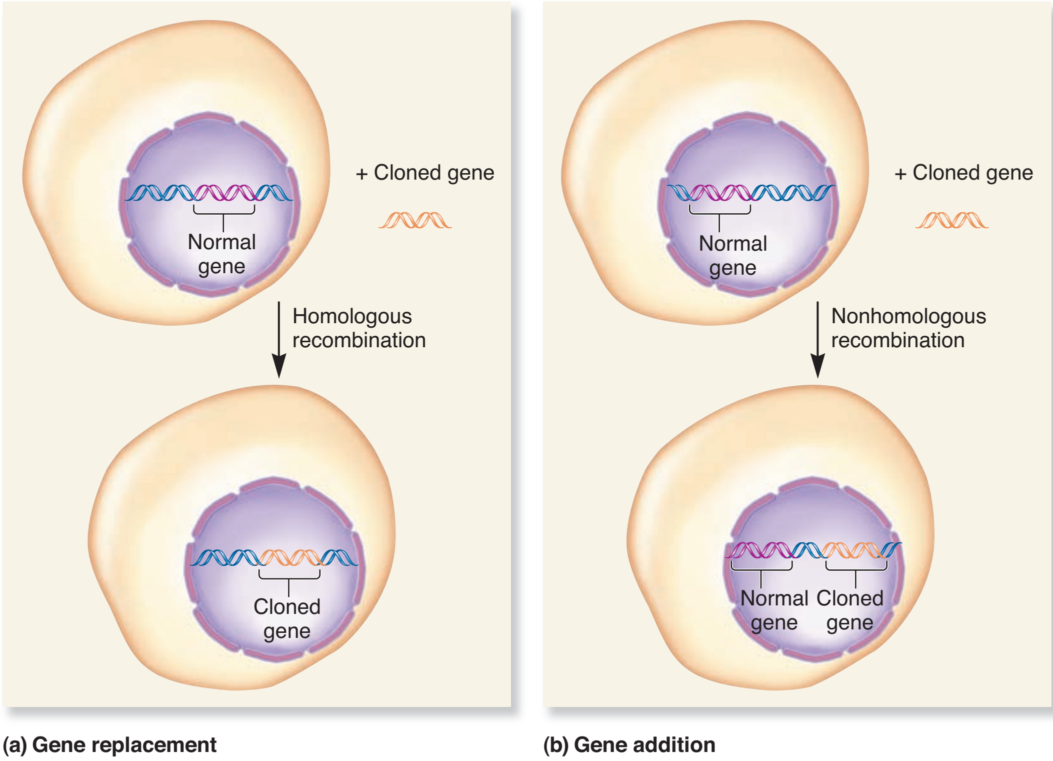
In Chapter 19, we considered methods of cloning genes. A common approach is to insert a chromosomal gene into a vector and then propagate the vector in living microorganisms such as bacteria or yeast cells. Cloned genes can also be introduced into animal and plant cells. However, to be inherited stably from generation to generation, the cloned gene must become integrated into one (or more) of the chromosomes that reside in the cell nucleus. The integration of cloned DNA into a chromosome occurs by recombination, which is described in Chapter 13.

Figure 20.3 illustrates how a cloned gene can integrate into a chromosome by recombination. If the genome of the host cell carries the same type of gene and if the cloned gene is swapped with the normal chromosomal gene by homologous recombination, then the cloned gene will replace the normal gene within the chromosome (Figure 20.3a). This process is termed **gene replacement.** If the cloned gene has been rendered inactive by mutation and replaces the normal gene, researchers can study how the loss of normal gene function affects the organism. This is called a **gene knockout.** More commonly, a cloned gene may be introduced into a cell and become integrated into the genome by nonhomologous recombination in a process known as **gene addition** (Figure 20.3b). Both the cloned and normal genes are present following gene addition.

Researchers may also introduce a cloned gene that is not already present in a particular species' genome. An interesting example of this type of gene addition involves the production of aquarium fish that "glow," the aptly named GloFish. The GloFish is a brand of transgenic zebrafish (*Danio rerio*) that fluoresce with a bright green, red, or yellow color (**Figure 20.4**). How were these fish produced? In 1999, Zhiyuan Gong and his colleagues started with a gene from jellyfish that encodes a green fluorescent protein and inserted it into the zebrafish genome by gene addition, causing the zebrafish to glow green. By placing the gene next to a gene promoter that would turn the gene on in the presence of certain environmental toxins, their goal was to eventually develop a fish that could be used to detect water pollution. The researchers subsequently collaborated with a company to market the fish for aquarium use. They developed a red fluorescent zebrafish by adding a gene from a sea coral, and a yellow fluorescent zebrafish by adding a variant of the jellyfish gene. In 2003, the GloFish became the first genetically modified organism to be sold as a pet. GloFish have been successfully marketed in several countries, including

FIGURE 20.3 **The introduction of a cloned gene into a cell can lead to gene replacement or gene addition.** In these two examples, the cloned gene is similar to a normal gene already present in the genome of a cell. **(a)** When the cloned gene undergoes homologous recombination and replaces the normal gene, this event is called gene replacement. **(b)** Alternatively, the cloned gene may recombine nonhomologously at some other chromosomal location, leading to gene addition.

Concept check: What is the difference between gene addition and gene replacement?



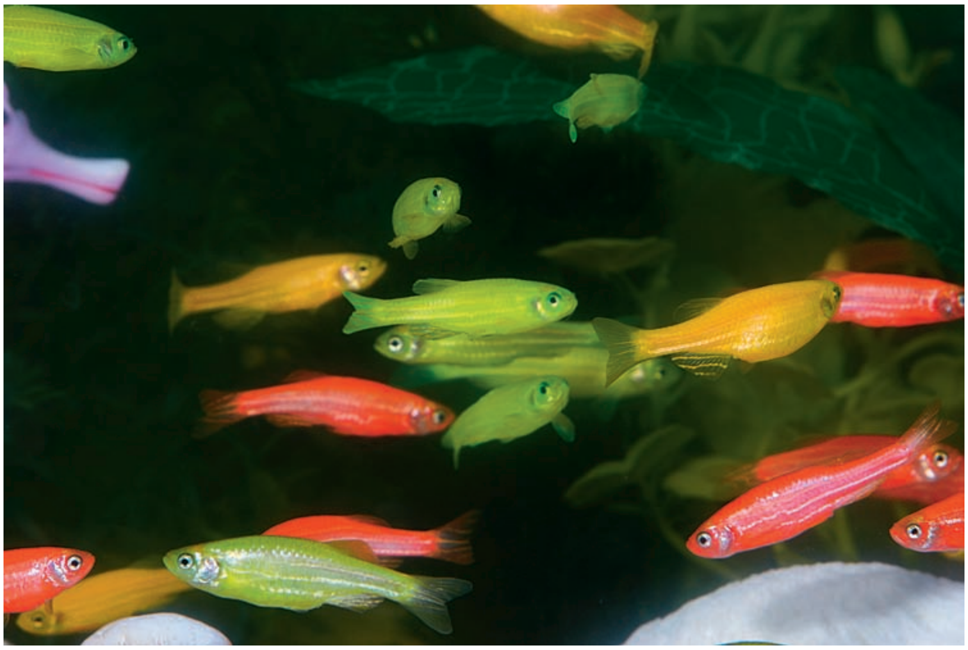


FIGURE 20.4 **The use of gene addition to produce fish that glow.** The aquarium fish shown here, named GloFish, are transgenic organisms that have received a gene from jellyfish or sea corals that encodes a fluorescent protein, causing them to glow green, red, or yellow.

the United States, although the GloFish is not sold in California, which has a ban on all genetically modified fish.

Molecular Biologists Can Produce Mice That Contain Gene Replacements

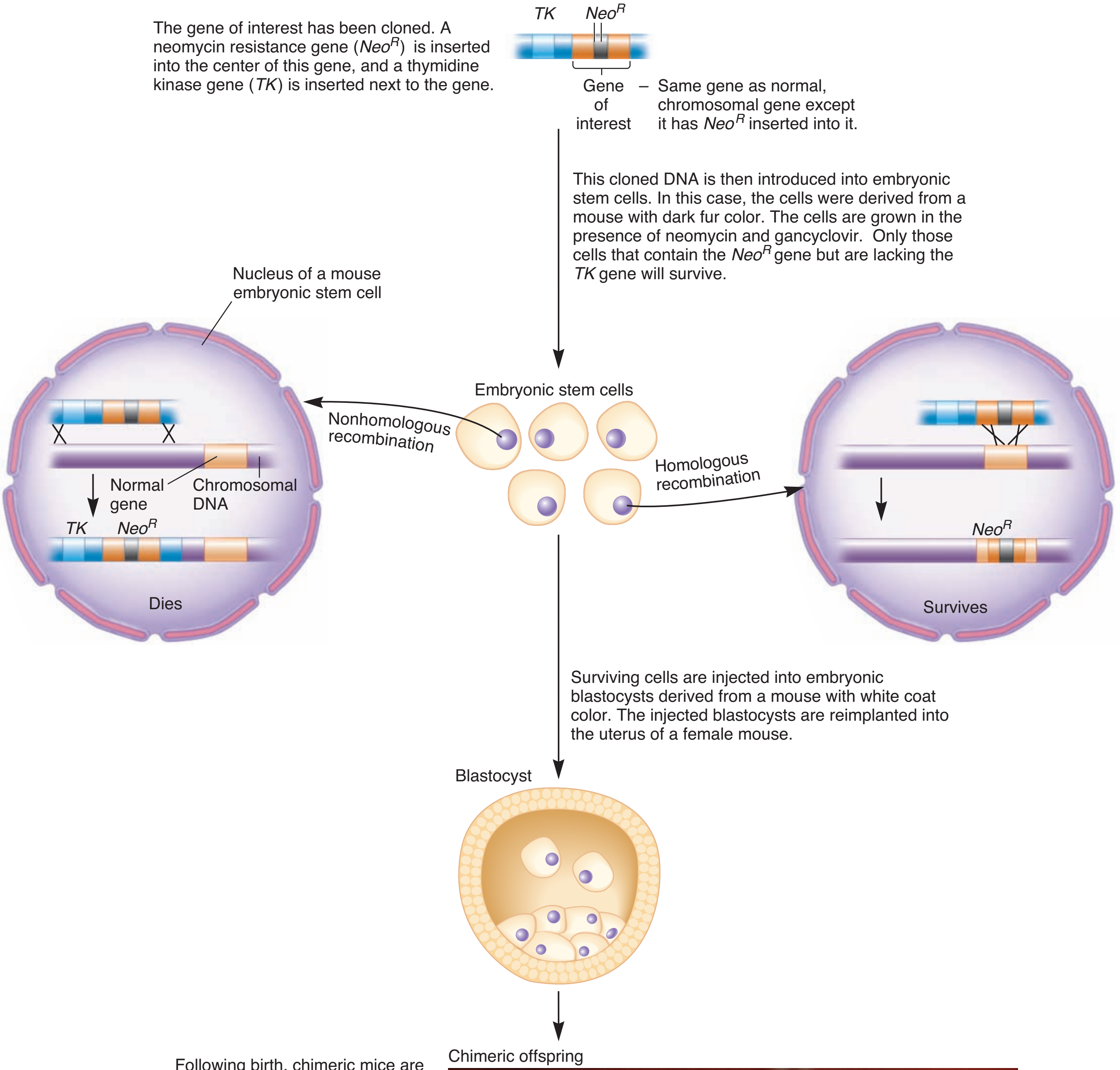
In bacteria and yeast, which have relatively small genomes, homologous recombination between cloned genes and the host cell chromosome occurs at a relatively high rate, so gene replacement is commonly achieved. Gene replacement is useful when a researcher or biotechnologist wants to make a gene mutation in the laboratory and then introduce the mutant gene into a living organism to compare the effects of the normal and mutant genes on the phenotype. However, in more complex eukaryotes with very large genomes, the introduction of cloned genes into cells is much more likely to result in gene addition rather than gene replacement. For example, when a cloned gene is introduced into a mouse cell, it undergoes homologous recombination less than 0.1% of the time. Most of the time (over 99.9%), gene addition occurs.

To produce mice with gene replacements, molecular biologists have devised laboratory procedures to preferentially select cells in which homologous recombination has occurred. One approach is shown in **Figure 20.5**. The gene of interest is found in a mouse chromosome and is cloned so it can be manipulated in vitro (see top of Figure 20.5). The cloning procedure involves two selectable marker genes that influence whether or not mouse embryonic cells can grow in the presence of certain drugs. First, the gene of interest is inactivated by inserting a neomycin-resistance gene (called *Neo*^{R}) into the center of its coding sequence. *Neo*^{R} provides cells with resistance to neomycin. Next, a thymidine kinase gene, designated *TK*, is inserted adjacent to the gene of interest but not within the gene itself. The *TK* gene renders cells sensitive to killing by a drug called gancyclovir.

After the cloned gene has been modified in vitro, it is introduced into mouse embryonic stem cells. (Stem cells are discussed later in this chapter.) How do researchers identify the cells in which homologous recombination has occurred? When the cells are grown in the presence of neomycin and gancyclovir, most nonhomologous recombinants are killed because they also carry the *TK* gene. In contrast, homologous recombinants, in which the normal gene has been partially replaced with the cloned gene, contain only the *Neo*^{R} gene, so they are resistant to both drugs. The surviving embryonic cells can then be injected into blastocysts, early embryos that are obtained from a pregnant mouse. In the example shown in Figure 20.5, the embryonic cells are from a mouse with dark fur, and the blastocysts are from a mouse with white fur. The embryonic cells can mix with the blastocyst cells to create a **chimera,** an organism that contains cells from two different individuals. To identify chimeras, the injected blastocysts are reimplanted into the uterus of a female mouse and allowed to develop. When this mouse gives birth, chimeras are easily identified because they contain patches of white and dark fur (see Figure 20.5).

Chimeric animals that contain a single-gene replacement can then be mated to other mice to produce offspring that carry the mutant gene. Because mice are diploid, researchers must make two or more subsequent crosses to produce a strain of mice that contains both copies of the mutant target gene. However, homozygous strains for a gene knockout cannot be produced if the mutant gene is lethal in the homozygous state.

An alternative method for producing genetically modified mice is to inject the desired gene into a fertilized egg. To conduct this type of experiment, researchers obtain mouse eggs and fertilize them in vitro. Immediately following fertilization, the cloned gene is injected into the sperm pronucleus—the haploid nucleus that has not yet fused with the egg nucleus. The cloned DNA then integrates into the genome, and the two pronuclei fuse to form



The gene of interest has been cloned. A neomycin resistance gene (*Neo*^{R}) is inserted into the center of this gene, and a thymidine kinase gene (*TK*) is inserted next to the gene.

Gene of interest – Same gene as normal, chromosomal gene except it has *Neo*^{R} inserted into it.

This cloned DNA is then introduced into embryonic stem cells. In this case, the cells were derived from a mouse with dark fur color. The cells are grown in the presence of neomycin and gancyclovir. Only those cells that contain the *Neo*^{R} gene but are lacking the *TK* gene will survive.

Surviving cells are injected into embryonic blastocysts derived from a mouse with white coat color. The injected blastocysts are reimplanted into the uterus of a female mouse.

TK
Neo^R
Nucleus of a mouse embryonic stem cell
Embryonic stem cells
Nonhomologous recombination
Homologous recombination
Normal gene
Chromosomal DNA
TK
Neo^R
Dies
Neo^R
Survives
Blastocyst
Chimeric offspring

Following birth, chimeric mice are identified as those that contain a coat with both dark and white fur. The appropriate crosses are made in order to produce mice that have two copies of the target gene.

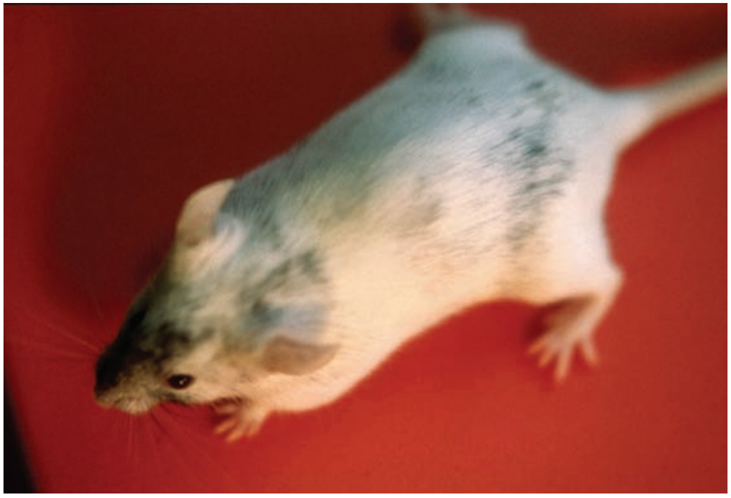


FIGURE 20.5 **Producing a gene replacement in mice.** The bottom of this figure shows a photograph of a chimeric mouse. Note the patches of black and white fur.

Concept check: *What is the purpose of the* Neo^R *gene?*

the diploid nucleus of a zygote. The zygote begins to divide and is introduced into the uterus of a female mouse, where it becomes implanted and grows. As discussed next, genetically modified mice are used in basic research and to study human diseases.

Gene Knockouts and Knockins Are Produced in Mice to Understand Gene Function and Human Disease

As we have seen, researchers can replace a normal mouse gene with one that has been inactivated by the insertion of an antibiotic-resistance gene. As mentioned, when a mouse is homozygous for an inactivated gene, this condition is called a gene knockout. The inactive mutant gene has replaced both copies of the normal gene. In other words, the function of the normal gene has been "knocked out." By creating gene knockouts, researchers can study how the loss of normal gene function affects the organism. Gene knockouts frequently have specific effects on the phenotype of a mouse, which helps researchers to determine that the function of a gene is critical within a particular tissue or during a specific stage of development. In many cases, however, a gene knockout produces no obvious phenotypic effect. One explanation is that a single gene may make such a small contribution to an organism's phenotype that its loss may be difficult to detect. Alternatively, another possible explanation for a lack of observable phenotypic change in a knockout mouse may involve **gene redundancy.** This means that when one type of gene is inactivated, another gene with a similar function may be able to compensate for the inactive gene.

A particularly exciting avenue of gene knockout research is its application in the study and treatment of human disease. How is this useful? Because humans share many genes with mice, observing the characteristics of knockout mice gives researchers information that can be used to better understand how a similar gene mutation may cause or contribute to a disease in humans. Examples of research areas in which knockout mice have been useful include cancer, obesity, heart disease, diabetes, and many inherited disorders. The use of knockouts in the area of functional genomics is also discussed in Chapter 22.

In contrast to knockouts, researchers may introduce genes into the mouse genome to study the effects of gene overexpression or to examine the effects of particular alleles, such as those that may cause disease in humans. To accomplish this, researchers can produce a **gene knockin** in which a gene of interest is inserted into a particular site in the mouse genome (**Figure 20.6**). In this example, the cloned gene is inserted into the middle of a segment of DNA from a noncritical site in the mouse genome. The noncritical segment is very long, which allows the transgene to be targeted to that specific, noncritical integration site by homologous recombination after it is introduced into mouse cells. Such gene knockins tend to result in a more consistent level of expression of the transgene compared with gene additions that may occur randomly in another place in the genome. Also, because a targeted transgene is not interfering with a critical mouse locus, the researcher can be more certain that any resulting phenotypic effect is due to the expression of the transgene.

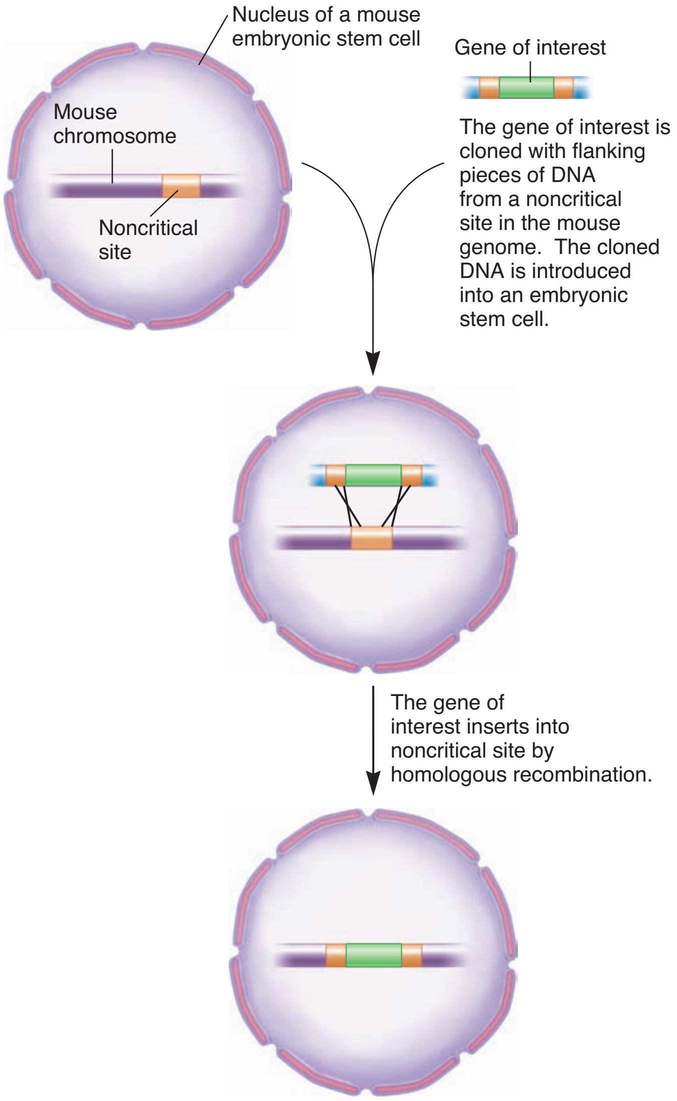

Nucleus of a mouse embryonic stem cell
Gene of interest
Mouse chromosome
Noncritical site
The gene of interest is cloned with flanking pieces of DNA from a noncritical site in the mouse genome. The cloned DNA is introduced into an embryonic stem cell.
The gene of interest inserts into noncritical site by homologous recombination.

FIGURE 20.6 **Producing a gene knockin in mice.** This simplified diagram illustrates the strategy of producing a gene knockin. The gene of interest is cloned, and the cloned gene is flanked by DNA sequences that are homologous to a noncritical site in the mouse genome. When introduced into a mouse embryonic stem cell, the gene inserts into the genome by homologous recombination, which is an example of a targeted gene addition.

Concept check: What is the difference between a gene knockout versus a gene knockin?

To study human diseases, researchers have produced strains of transgenic mice that harbor both gene knockouts and gene knockins. A strain of mice engineered to carry a mutation that is analogous to a mutation in a human gene is termed a **mouse model.** As an example, let's consider sickle cell disease, which is due to a mutation in the human β-globin gene (refer back to Figure 5.7). This gene encodes a polypeptide called β globin; adult hemoglobin is composed of both α-globin and β-globin polypeptides. When researchers produced a gene knockin by introducing the mutant human β-globin gene into mice, the resulting mice showed only mild symptoms of the disease, because they were expressing significant amounts of normal mouse hemoglobin. However, Chris Pászty and Edward Rubin produced a mouse model with multiple gene knockins and gene knockouts. In particular, the mice had gene knockins for the normal human α-globin gene and the mutant human β-globin gene from patients with sickle cell disease. The strain also had gene knockouts of the mouse α-globin gene and mouse β-globin gene:

normal human α-globin gene knockin;
mutant human β-globin gene knockin;
mouse α-globin gene knockout;
mouse β-globin gene knockout.

Therefore, these mice made adult hemoglobin just like people with sickle cell disease, but they did not produce any normal mouse hemoglobin. These transgenic mice exhibit the major features of sickle cell disease—sickled red blood cells, anemia, and multiorgan pathology. They have been useful as a model for studying the disease and testing potential therapies.

Biotechnology Holds Promise in Producing Transgenic Livestock

The technology for producing transgenic mice has been extended to other animals, and much research is under way to develop transgenic species of livestock, including fish, sheep, pigs, goats, and cattle. For some farmers, the ability to modify the characteristics of livestock via the introduction of cloned genes is an exciting prospect. In addition, work is currently under way to produce genetically modified pigs that are expected to be resistant to rejection mechanisms that occur following organ transplantation to humans. These strains may in time become a source of organs or cells for patients.

A novel avenue of research involves the production of medically important proteins in the mammary glands of livestock. This approach is sometimes called **molecular pharming.** (The term is also used to describe the manufacture of medical products by agricultural plants.) As shown in Table 20.3, several human proteins have been successfully produced in the milk of domestic livestock. Compared with the production of proteins in bacteria, one advantage is that certain proteins are more likely to function properly when expressed in mammals. This may be due to covalent modifications, such as the attachment of carbohydrate groups, which occur in eukaryotes but not in bacteria. In addition, certain proteins may be degraded rapidly or folded improperly when expressed in bacteria. Furthermore, the yield of recombinant proteins in milk can be quite large. Dairy cows, for example, produce about 10,000 liters of milk per year per cow. In some cases, a transgenic cow can produce approximately 1 g of the transgenic protein per liter of milk.

To introduce a human gene into an animal so the encoded protein will be secreted into its milk, the strategy is to insert the gene next to a milk-specific promoter. Eukaryotic genes often are expressed in a tissue-specific fashion. In mammals, certain genes are expressed specifically within the mammary gland so their protein product is secreted into the milk. Examples of milk-specific genes include genes that encode milk proteins such as β-lactoglobulin, casein, and whey acidic protein. To express a human gene that encodes a protein hormone into a domestic animal's milk, the promoter for a milk-specific gene is linked to the coding sequence for the human gene (Figure 20.7). The DNA is then injected into an oocyte, where it is integrated into the genome. The fertilized oocyte is then implanted into the uterus of a female animal, which later gives birth to a transgenic offspring. If the offspring is a female, the protein hormone encoded by the human gene is expressed within the mammary gland and secreted into the milk. The milk can then be obtained from the animal, and the human hormone isolated.

TABLE **20.3**

Proteins That Can Be Produced in the Milk of Domestic Animals

Protein	Host	Use
Lactoferrin	Cattle	Used as an iron supplement in infant formula
Tissue plasminogen activator (TPA)	Goat	Dissolves blood clots
Antibodies	Cattle	Used to combat specific infectious diseases
α_1-Antitrypsin	Sheep	Treatment of emphysema
Factor IX	Sheep	Treatment of certain inherited forms of hemophilia
Insulin-like growth factor	Cattle	Treatment of diabetes

20.2 REVIEWING THE KEY CONCEPTS

- A transgenic organism carries a gene from another species (see Figure 20.2).
- The integration of a cloned gene into a chromosome may result in gene replacement or gene addition (see Figures 20.3, 20.4).
- Gene knockouts and knockins in mice are made to study gene function and to produce mouse models for human diseases (see Figures 20.5, 20.6).
- Transgenic livestock may produce human proteins, such as hormones, in their milk (see Table 20.3, Figure 20.7).

20.2 COMPREHENSION QUESTIONS

1. When a cloned gene integrates into a genome by nonhomologous recombination, the end result is
   a. gene addition.
   b. gene replacement.
   c. a gene knockout.
   d. both a and b.
2. What is a chimera?
   a. A transgenic organism
   b. An organism that contains cells from two different individuals
   c. An organism that is produced by a gene knockin
   d. None of the above

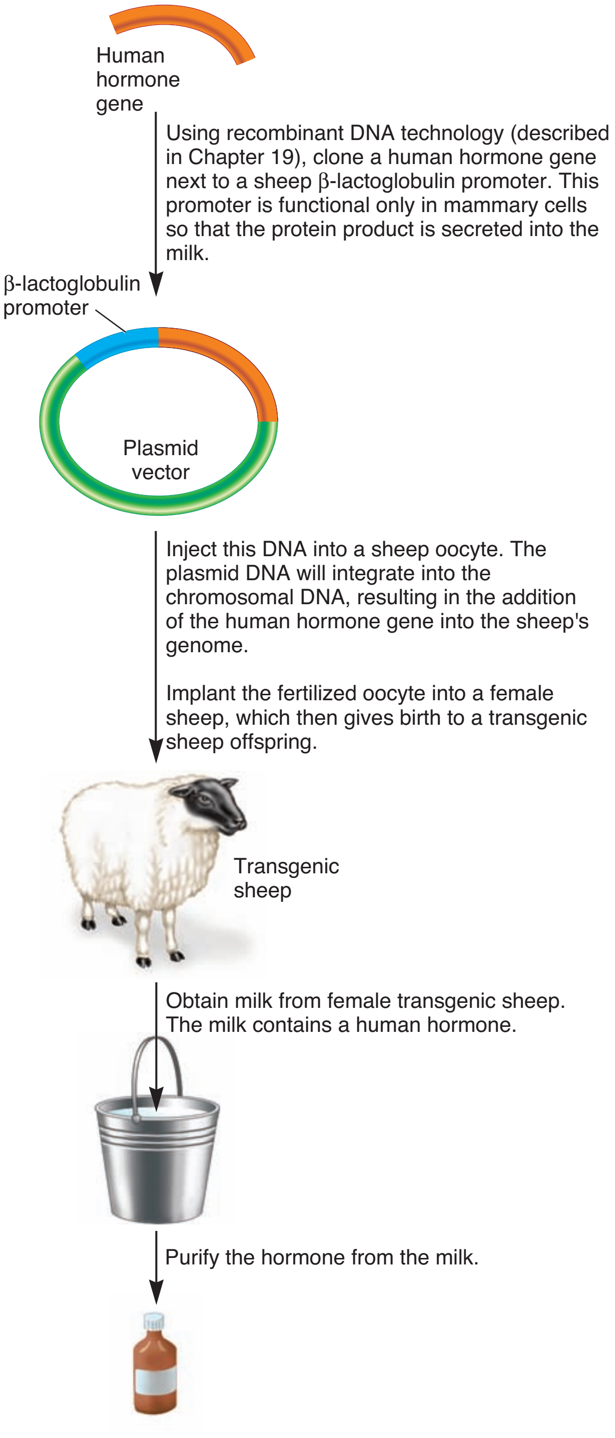

Human hormone gene
Using recombinant DNA technology (described in Chapter 19), clone a human hormone gene next to a sheep β-lactoglobulin promoter. This promoter is functional only in mammary cells so that the protein product is secreted into the milk.
β-lactoglobulin promoter
Plasmid vector
Inject this DNA into a sheep oocyte. The plasmid DNA will integrate into the chromosomal DNA, resulting in the addition of the human hormone gene into the sheep's genome.
Implant the fertilized oocyte into a female sheep, which then gives birth to a transgenic sheep offspring.
Transgenic sheep
Obtain milk from female transgenic sheep. The milk contains a human hormone.
Purify the hormone from the milk.

FIGURE 20.7 **Strategy for expressing human genes in a domestic animal's milk.** The β-lactoglobulin gene is normally expressed in mammary cells, whereas the human hormone gene is not. To express the human hormone gene in milk, the promoter from the milk-specific gene in sheep is linked to the coding sequence of the human hormone gene. In addition to the promoter, a short signal sequence may also be necessary so the protein is secreted from the mammary cells and into the milk.

Genes → Traits By using genetic engineering, researchers can give sheep the trait of producing a human hormone in their milk. This hormone can be purified from the milk and used to treat humans.

Concept check: *Why is a β-lactoglobulin promoter used?*

3. One strategy for producing a protein in the milk of a cow is to place the coding sequence of the gene of interest next to a __________ and then inject the gene into a __________.
   a. *lac* operon promoter, cow oocyte
   b. β-lactoglobulin promoter, cow oocyte
   c. *lac* operon promoter, cow mammary cell
   d. β-lactoglobulin promoter, cow mammary cell

20.3 REPRODUCTIVE CLONING AND STEM CELLS

The previous section focused on the area of biotechnology in which cloned genes are introduced into animals. Another aspect of biotechnology involves the cloning of whole organisms or the manipulation of stem cells. In this section, we will consider mammalian cloning and stem cell research. These topics have received enormous public attention due to the complex ethical issues they raise.

Researchers Have Succeeded in Cloning Mammals from Somatic Cells

The term *cloning* has different meanings. In Chapter 19, we discussed gene cloning, which involves methods that produce many copies of a gene. The cloning of an entire organism is a different matter. **Reproductive cloning** refers to methods that produce two or more genetically identical individuals. This happens occasionally in nature; identical twins are genetic clones that began from the same fertilized egg. Similarly, researchers can take mammalian embryos at an early stage of development (e.g., the two-cell to eight-cell stage), separate the cells, implant them into the uterus, and obtain multiple births of genetically identical individuals.

In the case of plants, cloning is an easier undertaking, as we will explore later in the chapter. Plants can be cloned from somatic cells. In most cases, it is relatively easy to take a cutting from a plant, expose it to growth hormones, and obtain a separate plant that is genetically identical to the original. However, this approach has not been possible with mammals. For several decades, scientists believed that chromosomes within the somatic cells of mammals had incurred irreversible genetic changes that render them unsuitable for cloning. However, this hypothesis has proven to be incorrect. In 1997, Ian Wilmut and his colleagues at the Roslin Institute in Scotland announced that a sheep, which they named Dolly, had been cloned using the genetic material from somatic cells.

How was Dolly produced? As shown in **Figure 20.8**, the researchers removed mammary cells from an adult female sheep and grew them in the laboratory. The researchers then extracted the nucleus from an egg cell of a different sheep and used electrical pulses to fuse the diploid mammary cell with the enucleated egg cell. After fusion, the zygote began embryonic development, and the resulting embryo was implanted into the uterus of a surrogate mother sheep. One hundred and forty-eight days later, Dolly was born.

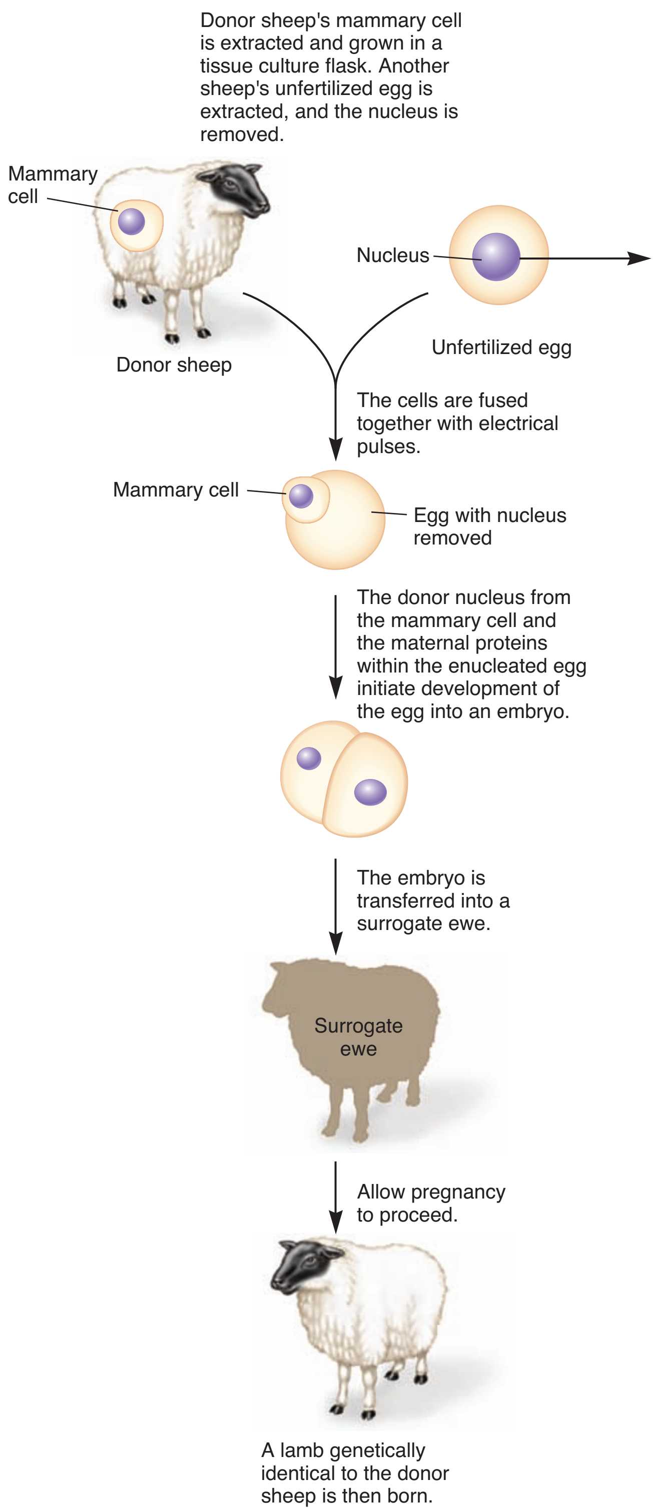

Donor sheep's mammary cell is extracted and grown in a tissue culture flask. Another sheep's unfertilized egg is extracted, and the nucleus is removed.
Mammary cell
Nucleus
Donor sheep
Unfertilized egg
The cells are fused together with electrical pulses.
Mammary cell
Egg with nucleus removed
The donor nucleus from the mammary cell and the maternal proteins within the enucleated egg initiate development of the egg into an embryo.
The embryo is transferred into a surrogate ewe.
Surrogate ewe
Allow pregnancy to proceed.
A lamb genetically identical to the donor sheep is then born.

FIGURE 20.8 **Protocol for the successful cloning of a sheep.**

Genes → Traits Dolly was (almost) genetically identical to the sheep that donated a mammary cell to create her. Dolly and the donor sheep were (almost) genetically identical in the same way that identical twins are; they carried the same set of genes and looked remarkably similar. However, they may have had minor genetic differences due to possible variation in their mitochondrial DNA and may have exhibited some phenotypic differences due to maternal effect or imprinted genes.

Concept check: *In the protocol, why is the nucleus of the oocyte removed?*

Although Dolly was clearly a clone of the donor female sheep, tests conducted when she was 3 years old suggested that she was "genetically older" than her actual age indicated. As mammals age, chromosomes in somatic cells tend to shorten from the telomeres—the ends of eukaryotic chromosomes. Therefore, older individuals have shorter chromosomes in their somatic cells than younger ones do. This shortening does not seem to occur in the cells of the germ line, however. When researchers analyzed the chromosomes in Dolly's somatic cells when she was about 3 years old, the lengths of her chromosomes were consistent with a sheep that was significantly older, say, 9 or 10 years old. The sheep that donated the somatic cell that produced Dolly was 6 years old, and her mammary cells had been grown in culture for several cell doublings before a mammary cell was fused with an oocyte. This led researchers to postulate that Dolly's shorter telomeres were a result of chromosome shortening in the somatic cells of the sheep that donated the nucleus. In 2003, the Roslin Institute announced the decision to euthanize 6-year-old Dolly after an examination showed progressive lung disease. Her death has raised concerns among scientists that the techniques used to produce Dolly could have caused premature aging.

With regard to telomere length, research in mice and cattle has shown different results; the telomeres of these cloned animals appear to be the correct length. For example, cloning was conducted on mice via the method described in Figure 20.8 for six consecutive generations. The cloned mice of the sixth generation had normal telomeres. Further research is necessary to determine if cloning via somatic cells has an effect on the length of telomeres in subsequent generations. However, other studies in mice point to various types of genetic flaws in cloned animals. For example, Rudolf Jaenisch and his colleagues used DNA microarray technology (described in Chapter 22) to analyze the transcription patterns of over 10,000 genes in cloned mice. As much as 4% of those genes were not expressed normally. Furthermore, research has shown that cloned mice die at a younger age than their naturally bred counterparts.

Mammalian cloning is still at an early stage of development. Nevertheless, the breakthrough of creating Dolly has shown that it is technically possible. In recent years, cloning from somatic cells has been achieved in several other mammalian species, including cattle, mice, goats, and pigs. In 2002, the first pet was cloned. She was named Carbon Copy, also called Copy Cat (**Figure 20.9**). Mammalian cloning may potentially have many practical applications. Cloning livestock would enable farmers to use the somatic cells from their best individuals to create genetically homogeneous herds, which could increase agricultural yield. However, such a genetically homogeneous herd may be more susceptible to rare diseases.

Though some people are concerned about the practical uses of cloning agricultural species, a majority have become very concerned with the possibility of human cloning. This prospect has raised a host of serious ethical questions. For example, some people feel that it is morally wrong and threatens the basic fabric of parenthood and family. Others feel that it is a technology that could offer a new avenue for reproduction, one that could be offered to infertile couples, for example. In the public sector,

FIGURE 20.9 **Carbon Copy, the first cloned pet.** The animal shown here was produced using a procedure similar to the one shown in Figure 20.8.

Concept check: *Is Carbon Copy a transgenic animal?*

the sentiment toward human cloning has been generally negative. Many countries have issued an all-out ban on human cloning, but others permit limited research in this area. Because the technology for cloning exists, our society will continue to wrestle with the legal and ethical aspects of cloning as it applies not only to animals but also to people.

Stem Cells Have the Ability to Divide and Differentiate into Different Cell Types

Stem cells supply the cells that construct our bodies from a fertilized egg. In adults, stem cells also replenish worn-out or damaged cells. To accomplish this task, stem cells have two common characteristics. First, they have the capacity to divide, and second, they can differentiate into one or more specialized cell types. As shown in Figure 20.10, the two daughter cells produced from the division of a stem cell can have different fates. One of the cells may remain an undifferentiated stem cell, while the other daughter cell can differentiate into a specialized cell type. With this type of asymmetrical division/differentiation pattern, the population of stem cells remains relatively constant, while providing a population of specialized cells. In the adult, this type of mechanism is needed to replenish cells that have a finite life span, such as skin epithelial cells and red blood cells.

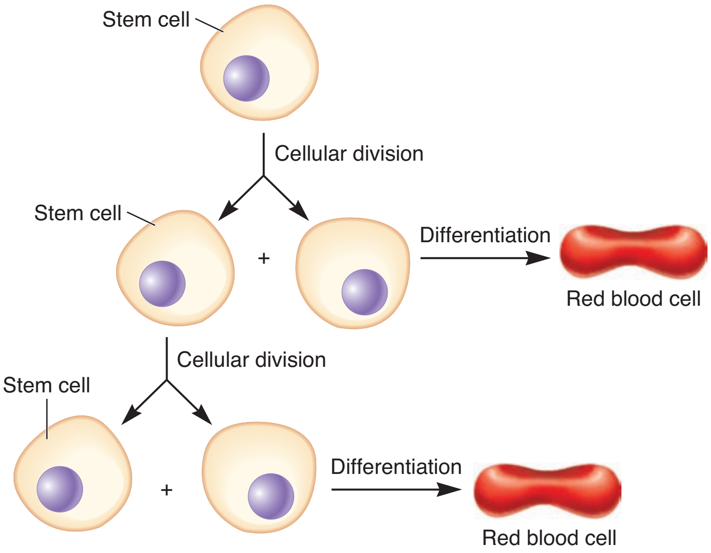


FIGURE 20.10 **Growth pattern of stem cells.** The two main traits that stem cells exhibit are an ability to divide and an ability to differentiate. When a stem cell divides, one of the two cells remains a stem cell, and the other daughter cell differentiates into a specialized cell type.

Concept check: *Explain why stem cells are not depleted during the life of an organism.*

In mammals, stem cells are commonly categorized according to their developmental stage and their ability to differentiate (Figure 20.11). The ultimate stem cell is the fertilized egg, which, via multiple cellular divisions, can give rise to an entire organism. A fertilized egg is considered **totipotent,** because it can give rise to all of the cell types in the adult organism. The early mammalian embryo contains **embryonic stem cells (ES cells),** which are found in the inner cell mass of the blastocyst. The blastocyst is the stage of embryonic development prior to uterine implantation—the preimplantation embryo. ES cells are **pluripotent,** which means they can differentiate into almost every cell type of the body. However, a single ES cell has lost the ability to produce an entire, intact individual. During the early fetal stage of development, the germ-line cells found in the gonads also are pluripotent. These cells are called **embryonic germ cells (EG cells).**

As mentioned, adults also contain stem cells, but these are thought to be multipotent or unipotent. A **multipotent** stem cell can differentiate into several cell types but far fewer than an ES cell. For example, hematopoietic stem cells (HSCs) found in the bone marrow can supply cells that populate two different tissues, namely, the blood and lymphoid tissues (Figure 20.12). Furthermore, each of these tissues contains several cell types. Multipotent HSCs can follow a pathway in which cell division produces a myeloid progenitor cell, which can then differentiate into a red blood cell, megakaryocyte, basophil, monocyte, eosinophil, neutrophil, or dendritic cell. Alternatively, an HSC can follow a path in which it becomes a lymphoid progenitor cell, which then differentiates into a T cell, B cell, natural killer cell, or dendritic cell. Other stem cells found in the adult seem to be **unipotent.** For example, primordial germ cells in the testis differentiate only into a single cell type, the sperm.

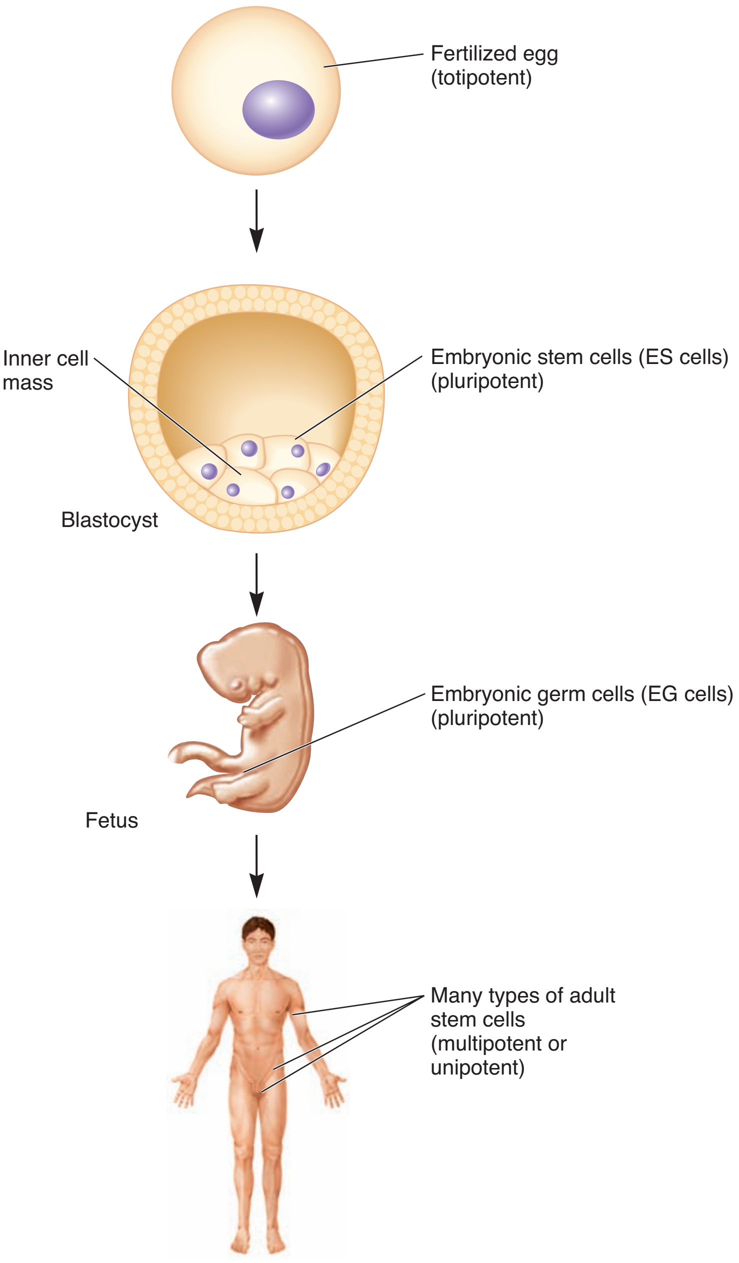

Fertilized egg
(totipotent)
Inner cell mass
Embryonic stem cells (ES cells)
(pluripotent)
Blastocyst
Embryonic germ cells (EG cells)
(pluripotent)
Fetus
Many types of adult stem cells
(multipotent or unipotent)

FIGURE 20.11 **Occurrence of stem cells at different stages of human development.**

Stem Cells Have the Potential to Treat a Variety of Diseases

What are the potential uses of stem cells? Interest in stem cells centers on two main areas. Because stem cells have the capacity to differentiate into multiple cell types, the study of stem cells may help us to understand basic genetic mechanisms that underlie the process of development, the details of which are described in Chapter 24. A second compelling reason why people have become interested in stem cells is their potential to treat human diseases or injuries that cause cell and tissue damage. This application has already become a reality in certain cases. For example, bone marrow transplantation is used to treat patients with certain forms of cancers. Such patients may be given radiation treatments that destroy their immune systems. When these patients are injected with bone marrow from a healthy person, the stem cells within the transplanted marrow have the ability to proliferate and differentiate within their bodies and provide them with a functioning immune system.

Renewed interest in the use of stem cells in the potential treatment of many other diseases was fostered in 1998 by studies of two separate teams, headed by James Thomson and John Gearhart, showing that embryonic cells, either ES or EG cells, can be successfully propagated in the laboratory. As mentioned, ES and EG cells are pluripotent and therefore have the capacity to produce many different kinds of tissue. As shown in **Table 20.4**, embryonic cells could potentially be used to treat a wide variety of diseases associated with cell and tissue damage. By comparison, it would be difficult, based on our modern knowledge, to treat these diseases with adult stem cells because of the inability to locate most types of adult stem cells within the body and successfully grow them in the laboratory. Even HSCs are elusive. In the bone marrow, about 1 cell in 10,000 is a stem cell, yet that is enough to populate all of the blood and lymphoid cells of the body. The stem cells of most other adult tissues are equally if not more difficult to locate. In addition, with the exception of stem cells in the blood, other types of stem cells in the adult body are difficult to remove in sufficient numbers for transplantation. By comparison, ES and EG cells are easy to identify and have the great advantage of rapid growth in the laboratory. For these reasons, ES and EG cells offer a greater potential for transplantation, based on our current knowledge of stem cell biology.

For ES or EG cells to be used in transplantation, researchers need to derive methods that cause them to differentiate into the appropriate type of tissue. For example, if the goal is to repair a spinal cord injury, ES or EG cells need the appropriate cues that cause them to differentiate into neural tissue. At present, much research is needed to understand and potentially control the fate of ES or EG cells. Currently, researchers speculate that a complex

TABLE **20.4**

Potential Uses of Stem Cells to Treat Diseases

Cell/Tissue Type	Disease Treatment
Neural	Implantation of cells into the brain to treat Parkinson disease
	Treatment of neurological injuries such as those to the spinal cord
Skin	Treatment of burns and other types of skin disorders
Cardiac	Repair of heart damage associated with heart attacks
Cartilage	Repair of joints damaged by injury or arthritis
Bone	Repair of damaged bone or replacement with new bone
Liver	Repair or replacement of liver tissue that has been damaged by injury or disease
Skeletal muscle	Repair or replacement of damaged muscle

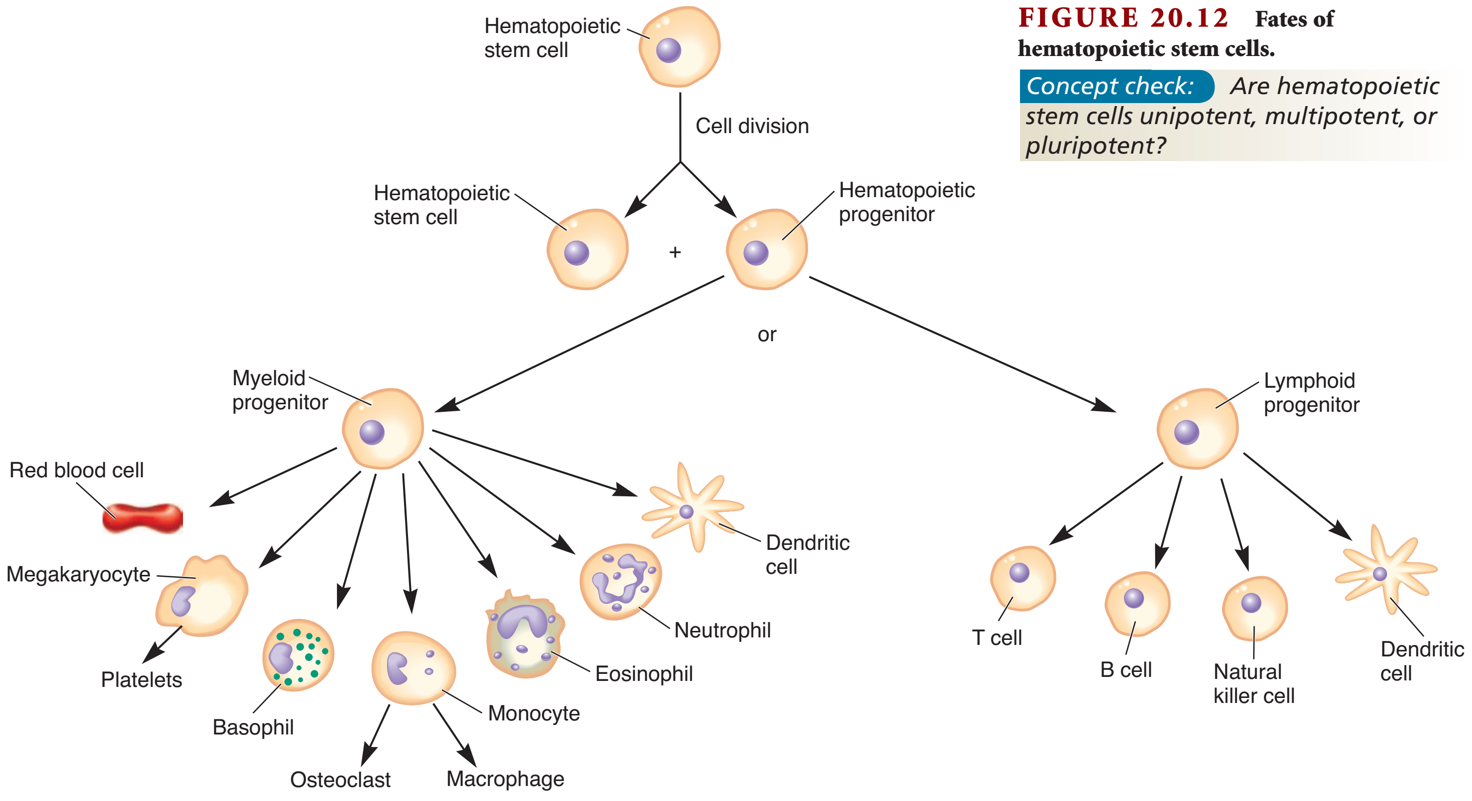


FIGURE 20.12 **Fates of hematopoietic stem cells.**

Concept check: Are hematopoietic stem cells unipotent, multipotent, or pluripotent?

variety of factors determine the developmental fates of stem cells. These include internal factors within the stem cells themselves, as well as external factors such as the properties of neighboring cells and the presence of hormones and growth factors in the environment.

From an ethical perspective, the primary issue that raises debate is the source of the stem cells for research and potential treatments. Most ES cells have been derived from human embryos that were produced from in vitro fertilization and were subsequently not used. Most EG cells are obtained from aborted fetuses. Some feel that it is morally wrong to use such tissue in research and/or the treatment of disease or they fear this technology could lead to intentional abortions for the sole purpose of obtaining fetal tissues for transplantation. Alternatively, others feel that the embryos and fetuses that provide the ES and EG cells are not going to become living individuals, and therefore, it is beneficial to study these cells and use them in a positive way to treat human diseases and injury. It is not clear whether these two opposing viewpoints can reach a common ground. As a compromise, some governments have enacted laws that limit or prohibit the use of embryos or fetuses to obtain stem cells, yet permit the use of stem cell lines that are already available in research laboratories.

If stem cells could be obtained from adult cells and propagated in the laboratory, an ethical dilemma may be avoided because most people do not have serious moral objections to current procedures such as bone marrow transplantation. In 2006, work by Shinya Yamanaka and colleagues showed that adult mouse fibroblasts (a type of connective tissue cell) could become pluripotent via the injection of four different genes that encode transcription factors. In 2007, Yamanaka's laboratory and two other research groups showed that such induced pluripotent stem cells (iPS) can differentiate into all cell types when injected into mouse blastocysts and grown into baby mice. Though further research is still needed, these recent results indicate that adult cells can be reprogrammed to become embryonic stem cells.

20.3 REVIEWING THE KEY CONCEPTS

- Mammalian reproductive cloning can be achieved by fusing somatic cells to oocytes with their nuclei removed (see Figures 20.8, 20.9).
- Stem cells have the ability to divide and differentiate. Stem cells may be totipotent, pluripotent, multipotent, or unipotent (see Figures 20.10–20.12).
- Stem cells have the potential to treat a variety of human diseases (see Table 20.4).

20.3 COMPREHENSION QUESTIONS

1. During mammalian reproductive cloning, __________ is fused with ___________.
   a. a somatic cell, a stem cell
   b. a somatic cell, an egg cell
   c. a somatic cell, an enucleated egg cell
   d. an enucleated somatic cell, an egg cell
2. Which of the following is a key feature of stem cells?
   a. They have the ability to divide.
   b. They have the ability to differentiate.
   c. They are always pluripotent.
   d. Both a and b

20.4 GENETICALLY MODIFIED PLANTS

As we have seen, researchers have succeeded in making genetically modified animals for a variety of reasons. In this section, we will examine the methods that scientists follow to make transgenic plants.

A large amount of research has been aimed at the use of transgenic species in agriculture. For centuries, agriculture has relied on selective breeding programs to produce crops and livestock with desirable characteristics. For agriculturally important species, this often means the production of strains that are larger, are disease-resistant, and yield high-quality food. Agricultural scientists can now complement traditional breeding strategies with modern molecular genetic approaches. In the mid-1990s, genetically modified crops first became commercialized. Since that time, their use has progressively increased. In 2009, roughly 25% of all agricultural crops were transgenic. Worldwide, more than 100 million hectares (247 million acres) of transgenic crops were planted. In this section, we will discuss some current and potential uses of transgenic plants in agriculture.

Agrobacterium tumefaciens and Other Methods Can Be Used to Make Transgenic Plants

As we have seen, the introduction of cloned genes into embryonic cells can produce transgenic animals. The production of transgenic plants is somewhat easier because some somatic cells are totipotent, which means they are capable of developing into an entire organism. Therefore, a transgenic plant can be made by the introduction of cloned genes into somatic tissue, such as the tissue of a leaf. After the cells of a leaf have become transgenic, an entire plant can be regenerated by the treatment of the leaf with plant growth hormones, which cause it to form roots and shoots.

Molecular biologists can use the bacterium *Agrobacterium tumefaciens*, which naturally infects plant cells, to produce transgenic plants. A plasmid from the bacterium, known as the Ti plasmid (Tumor-inducing plasmid), naturally induces tumor formation after a plant has been infected (**Figure 20.13a**). A segment of the plasmid DNA, known as **T DNA** (for transferred DNA), is transferred from the bacterium to the infected plant cells. The T DNA from the Ti plasmid is integrated into the chromosomal DNA of the plant cell by recombination. After this occurs, genes within the T DNA that encode plant growth hormones cause uncontrolled plant cell growth. This produces a cancerous plant growth known as a crown gall tumor (**Figure 20.13b**).

Because *A. tumefaciens* inserts its T DNA into the chromosomal DNA of plant cells, it can be used as a vector to introduce cloned genes into plants. Molecular geneticists have been able to modify the Ti plasmid to make this an efficient process. Such vectors are called **T-DNA vectors.** The T DNA genes that cause the development of a gall have been identified. Fortunately for genetic engineers, when these genes are deleted, the T DNA is still taken up into plant cells and integrated within the plant chromosomal DNA. However, a gall tumor does not form. In addition, geneticists have inserted selectable marker genes into the T DNA to identify plant cells that have taken up the T DNA. A gene that provides resistance to the antibiotic kanamycin (*kan*^{R}) is a commonly used selectable marker. The T-DNA vectors used in cloning experiments are also modified to contain unique restriction sites for the convenient insertion of any gene.

Figure 20.14 shows the general strategy for producing transgenic plants via T-DNA-mediated gene transfer. A gene of interest is inserted into a genetically engineered T-DNA vector and then transformed into *A. tumefaciens*. Plant cells are exposed to the transformed *A. tumefaciens*. After allowing time for infection, the plant cells are exposed to the antibiotics kanamycin and carbenicillin. Carbenicillin kills *A. tumefaciens*, and kanamycin kills any plant cells that have not taken up the T DNA with the antibiotic-resistance gene. Therefore, the only surviving cells are those plant cells that have integrated the T DNA into their genome. Because the T DNA also contains the cloned gene of interest, the selected plant cells are expected to have received this cloned gene as well. The cells are then transferred to a medium that contains the plant growth hormones necessary for the regeneration of entire plants. These plants can then be analyzed to verify that they are transgenic plants containing the cloned gene.

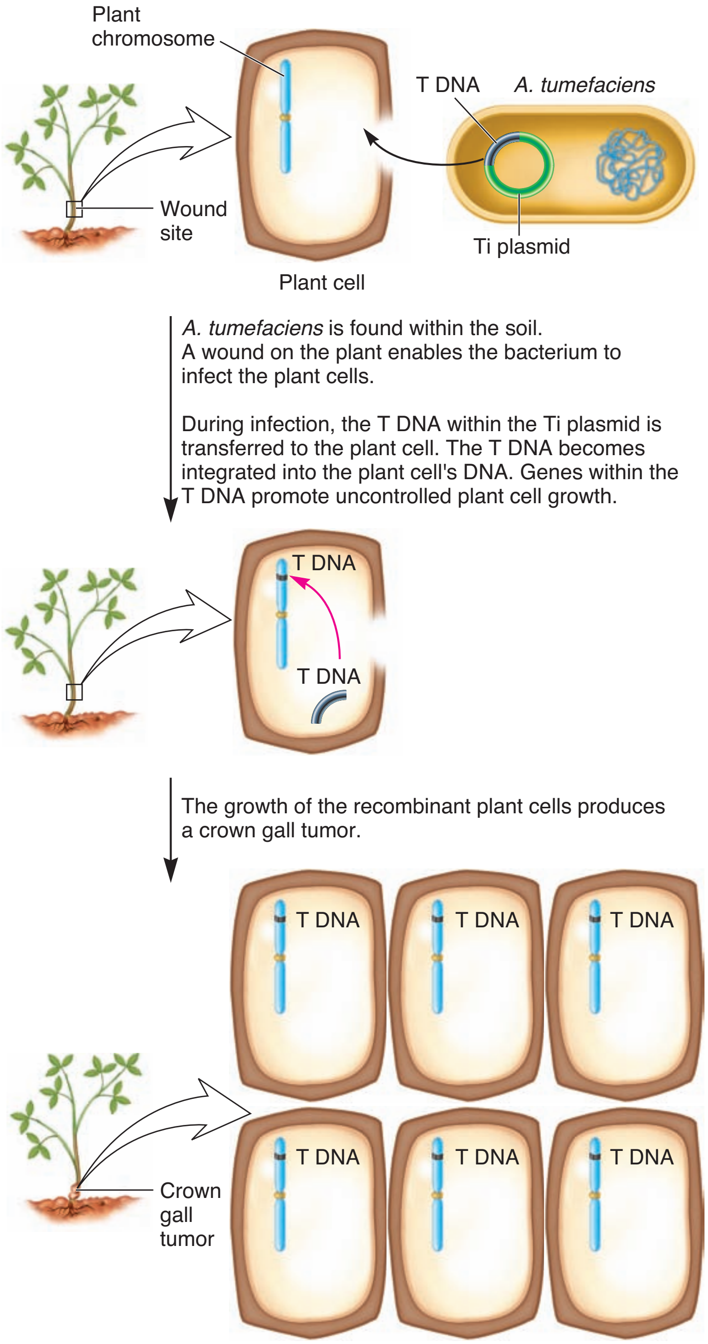
A. tumefaciens infects a wide range of plant species, including most dicotyledonous plants, most gymnosperms, and some monocotyledonous plants. However, not all plant species are infected by this bacterium. Fortunately, other methods are available for introducing genes into plant cells. Another common way to produce transgenic plants is an approach known as **biolistic gene transfer.** In this method, plant cells are bombarded with high-velocity microprojectiles coated with DNA. When fired on by this "gene gun," the microprojectiles penetrate the cell wall and membrane, thereby entering the plant cell. The cells that take up the DNA are identified with a selectable marker and regenerated into new plants.

DNA can also be introduced into plant cells by **microinjection**—the use of microscopic-sized needles—or by **electroporation**—the use of electrical current to create temporary pores in the plasma membrane. Because the rigid plant cell wall is a difficult barrier for DNA entry, another approach involves the use of **protoplasts,** which are plant cells with their cell walls removed. DNA can be introduced into protoplasts using a variety of methods, including treatment with polyethylene glycol and calcium phosphate.

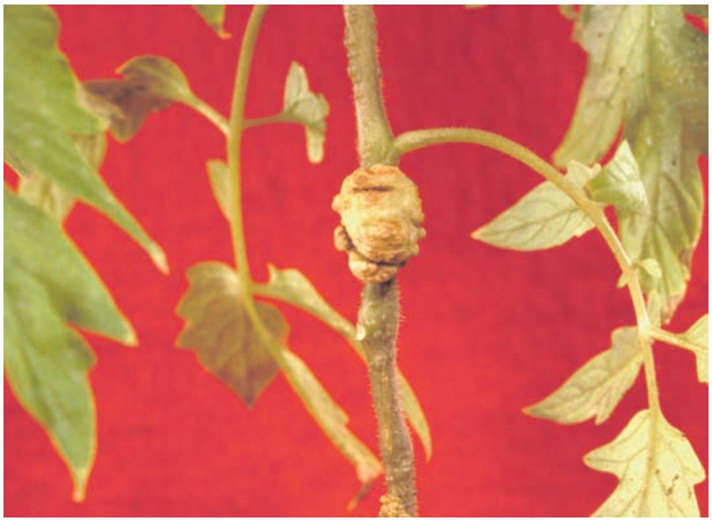
The production of transgenic plants has been achieved for many agriculturally important plant species, including alfalfa, corn, cotton, soybean, tobacco, and tomato. Some of the applications of transgenic plants are described next.

Transgenic Plants Can Be Given Characteristics That Are Agriculturally Useful

Various traits can be modified in transgenic plants (**Table 20.5**). Frequently, transgenic research has sought to produce plant strains resistant to insects, disease, and herbicides. For example, transgenic plants highly tolerant of particular herbicides have been made. The Monsanto Company has produced transgenic


Plant chromosome
T DNA
A. tumefaciens
Wound site
Ti plasmid
Plant cell
A. tumefaciens is found within the soil. A wound on the plant enables the bacterium to infect the plant cells.
During infection, the T DNA within the Ti plasmid is transferred to the plant cell. The T DNA becomes integrated into the plant cell's DNA. Genes within the T DNA promote uncontrolled plant cell growth.
T DNA
T DNA
The growth of the recombinant plant cells produces a crown gall tumor.
T DNA
T DNA
T DNA
T DNA
T DNA
T DNA
Crown gall tumor

(a) The production of a crown gall tumor by *A. tumefaciens* infection



(b) A crown gall tumor on a pecan tree

FIGURE 20.13 ***Agrobacterium tumefaciens* infecting a plant and causing a crown gall tumor.**

plant strains tolerant of glyphosate, the active agent in the herbicide Roundup. The herbicide remains effective against weeds, but the herbicide-resistant crop is spared (**Figure 20.15**).

Another important approach is to make plant strains that are disease-resistant. In many cases, virus-resistant plants have been developed by introducing a gene that encodes a viral coat protein. When the plant cells express the viral coat protein, they become resistant to infection by that pathogenic virus.

Many transgenic plants have been approved for human consumption. A successful example of the use of transgenic plants has involved the introduction of genes from *Bacillus thuringiensis* (Bt). As discussed earlier, this bacterium produces toxins that kill certain types of caterpillars and beetles and has been widely used as a biological control agent for several decades. These toxins are proteins encoded in the genome of *B. thuringiensis*. Researchers have succeeded in cloning toxin genes from *B. thuringiensis* and transferring those genes into plants. Such Bt varieties of plants produce the toxins themselves and therefore are resistant to many types of caterpillars and beetles. Examples of commercialized crops include Bt corn (**Figure 20.16a**) and Bt cotton. Since their introduction

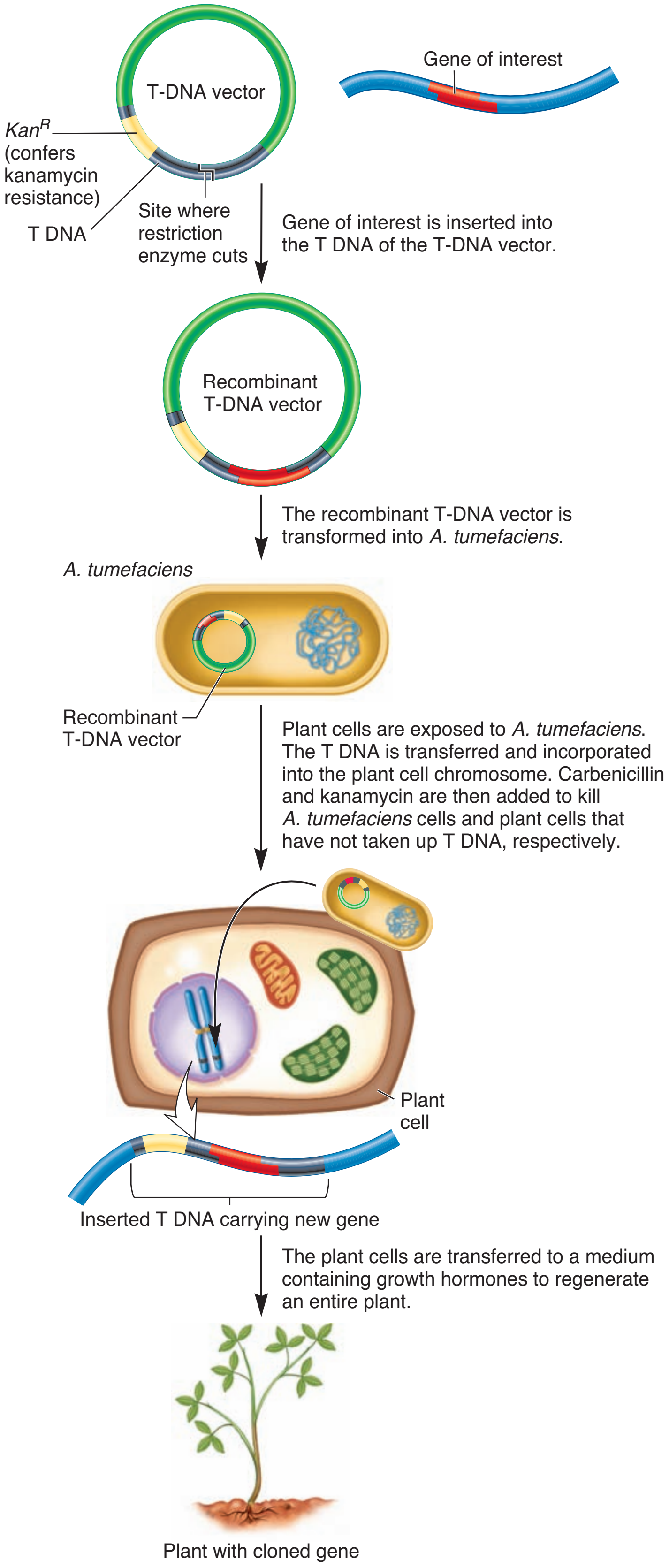

T-DNA vector
Gene of interest
Kan^R (confers kanamycin resistance)
T DNA
Site where restriction enzyme cuts
Gene of interest is inserted into the T DNA of the T-DNA vector.
Recombinant T-DNA vector
The recombinant T-DNA vector is transformed into *A. tumefaciens*.
A. tumefaciens
Recombinant T-DNA vector
Plant cells are exposed to *A. tumefaciens*. The T DNA is transferred and incorporated into the plant cell chromosome. Carbenicillin and kanamycin are then added to kill *A. tumefaciens* cells and plant cells that have not taken up T DNA, respectively.
Plant cell
Inserted T DNA carrying new gene
The plant cells are transferred to a medium containing growth hormones to regenerate an entire plant.
Plant with cloned gene

FIGURE 20.14 **The transfer of genes into plants using a T-DNA vector from** ***A. tumefaciens.***

Concept check: Which portion of a T-DNA vector is transferred to a plant?

TABLE 20.5
Traits That Have Been Modified in Transgenic Plants

Trait	Examples
Plant Protection	
Resistance to viral, bacterial, and fungal pathogens	Transgenic plants that express the pokeweed antiviral protein are resistant to a variety of viral pathogens.
Resistance to insects	Transgenic plants that express the CryIA protein from *Bacillus thuringiensis* are resistant to a variety of insects (see Figure 20.16).
Resistance to herbicides	Transgenic plants can express proteins that render them resistant to particular herbicides (see Figure 20.15).
Plant Quality	
Improvement in storage	Transgenic plants can express antisense RNA that silences a gene involved in fruit softening.
Plant composition	Transgenic strains of canola have been altered with regard to oil composition; the seeds of the Brazil nut have been rendered methionine-rich via transgenic technology.
New Products	
Biodegradable plastics	Transgenic plants have been made that can synthesize polyhydroxyalkanoates, which are used as biodegradable plastics.
Vaccines	Transgenic plants have been modified to produce vaccines in their leaves against many human and animal diseases, including hepatitis B, cholera, and malaria.
Pharmaceuticals	Transgenic plants have been made that produce a variety of medicines, including human interferon-α (to fight viral diseases and cancer), human epidermal growth factor (for wound repair), and human aprotinin (for reducing blood loss during transplantation surgery).
Antibodies	Human antibodies have been made in transgenic plants to battle various diseases such as non-Hodgkin lymphoma.

in 1996, the commercial use of these two Bt crops has steadily increased (**Figure 20.16b**).

The introduction of transgenic plants into agriculture has been strongly opposed by some people. What are the potential risks? One problem is that transgenes in commercial crops can endanger native species. For example, Bt crops may kill pollinators of native species. Another issue is that the planting of transgenic crops may lead to the proliferation of Bt-resistant insects. Though Bt-resistance is rare, it can occasionally happen. To minimize the chances of Bt resistance, researchers are producing transgenic strains that carry more than one toxin gene, which makes it more difficult for insect resistance to arise. Despite these and other concerns, many farmers are embracing transgenic crops, and their use continues to rise.

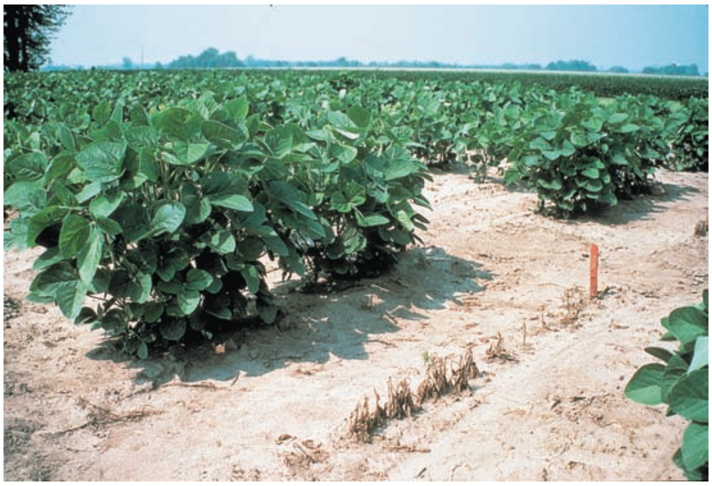


FIGURE 20.15 **Transgenic plants that are resistant to glyphosate.**

Genes → Traits This field of soybean plants has been treated with glyphosate. The plants on the left have been genetically engineered to contain a herbicide-resistance gene. They are resistant to killing by glyphosate. By comparison, the dead plants in the row with the orange stick do not contain this gene.

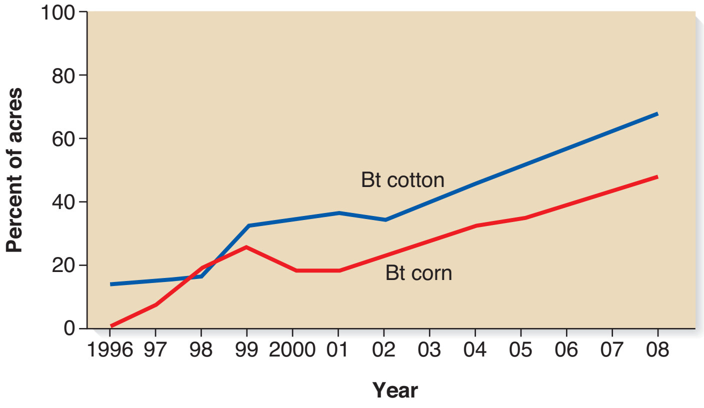
20.4 REVIEWING THE KEY CONCEPTS

- *Agrobacterium tumefaciens* transfers T DNA to plants. Researchers have used T-DNA vectors to make genetically modified plants (see Figures 20.13, 20.14).
- Researchers have made many transgenic plants that have traits that are useful to humans, including herbicide and pesticide resistance (see Table 20.5, Figures 20.15, 20.16).

20.4 COMPREHENSION QUESTIONS

1. When *A. tumefaciens* is used to make a transgenic plant, a gene of interest is inserted into a __________, which is first transformed into __________. The __________ is then transferred to a plant.
   a. viral vector, *E. coli*, gene of interest
   b. T-DNA vector, *A. tumefaciens*, T-DNA carrying the gene of interest
   c. T-DNA vector, *A. tumefaciens*, gene of interest alone
   d. T-DNA vector, *E. coli*, gene of interest alone
2. In addition to T-DNA vectors, other technologies to produce transgenic plants include
   a. biolistic gene transfer.
   b. microinjection.
   c. electroporation.
   d. all of the above.

(a) A field of Bt corn



(b) Bt corn and Bt cotton usage since 1996

FIGURE 20.16 **The production of Bt crops.** **(a)** A field of Bt corn. These corn plants carry an endotoxin gene from *Bacillus thuringiensis* that provides them with resistance to insects such as corn borers, which are a major pest of corn plants. **(b)** A graph showing the increase in usage of Bt corn and Bt cotton in the United States since their commercial introduction in 1996.

20.5 HUMAN GENE THERAPY

Throughout this textbook, we have considered examples in which mutant genes cause human diseases. In Chapter 23, we will discuss some examples that include rare inherited disorders and more common diseases such as cancer. Because mutant genes cause disease, geneticists are actively pursuing the goal of using normal, cloned genes to compensate for defects in mutant genes. **Gene therapy** is the introduction of cloned genes into living cells to treat disease. It is a potential method for treating a wide variety of illnesses.

Many current research efforts in gene therapy are aimed at alleviating inherited human diseases. Over 7000 human genetic diseases are known to involve a single gene abnormality. Familiar examples include cystic fibrosis, sickle cell disease, and hemophilia. In addition, gene therapies have also been aimed at treating diseases such as cancer and cardiovascular disease, which may occur later in life. Some scientists are even pursuing research that will use gene therapy to combat infectious diseases such as AIDS. Even though gene therapy is still at an early stage of development, a large amount of research has already been conducted. Unfortunately, success has been limited, and relatively few patients have been treated with gene therapy. Nevertheless, a few results have been somewhat promising. **Table 20.6**

TABLE **20.6**

Future Prospects in Gene Therapy

Type of Disease	Treatment of
Blood	Sickle cell disease, hemophilia, severe combined immunodeficiency disease (SCID)
Metabolic	Glycogen storage diseases, lysosomal storage diseases, phenylketonuria
Muscular	Duchenne muscular dystrophy, myotonic muscular dystrophy
Lung	Cystic fibrosis
Cancer	Brain tumors, breast cancer, colorectal cancer, malignant melanoma, ovarian cancer, and several other types of malignancies
Cardiovascular	Atherosclerosis, essential hypertension
Infectious	AIDS, possibly other viral diseases that involve latent infections

describes several types of diseases that are being investigated as potential targets for gene therapy. In this section, we will examine the approaches to gene therapy and how it may be used to treat human disease.

Gene Therapy Involves the Introduction of Cloned Genes into Human Cells

A key step in gene therapy is the introduction of a cloned gene into the cells of people. This is a challenging aspect of gene therapy because it requires that many cells of a person's body take up the cloned gene and express it. Otherwise, the disease symptoms will not be corrected. The techniques for transferring a cloned gene into human cells can be categorized as nonviral and viral gene transfer methods. The most common nonviral technique involves the use of **liposomes,** which are lipid vesicles (**Figure 20.17a**). The DNA containing the gene of interest is complexed with liposomes that carry a positive charge (i.e., cationic liposomes). The DNA-liposome complexes are taken into cells via endocytosis, in which a portion of the plasma membrane invaginates and creates an intracellular vesicle known as an endosome; the liposome is degraded within the endosome. The DNA is then released into the cytosol, imported into the nucleus, and then integrated into a chromosome of the target cell. An advantage of gene transfer via liposomes is that the liposomes do not elicit an immune response. A disadvantage is that the efficiency of gene transfer may be very low.

A second way to transfer genes into human cells is via viruses. Commonly used viruses for gene therapy include retroviruses, adenoviruses, and parvoviruses. The genetic modification of these viral genomes has led to the development of gene therapy vectors with a capacity to infect cells or tissues, much like the ability of wild-type viruses to infect cells. However, in contrast to wild-type viruses, gene therapy viral vectors have been genetically engineered so they can no longer replicate within target cells. Nevertheless, the genetically engineered viruses are naturally taken up by cells via endocytosis (**Figure 20.17b**). The viral coat disassembles, and the viral genome is released into the cytosol. In the case of retroviruses, the genome is RNA, which is reverse-transcribed into DNA. The viral DNA, which carries a gene of interest, travels into the nucleus and is then integrated into a chromosome of the target cell.

A key advantage of viral vectors is their ability to efficiently transfer cloned genes to a variety of human cell types. However, a major disadvantage of viral-mediated gene therapy is the potential to evoke an undesirable immune response when injected into a patient. The inflammatory responses induced by adenovirus particles, for example, can be very strong and even fatal. Therefore, much effort has been aimed at preventing the inflammatory responses mediated by virus particles. These include the application of immunosuppressive drugs and the generation of less immunogenic viral vectors by further genetic modification within the viral genome.

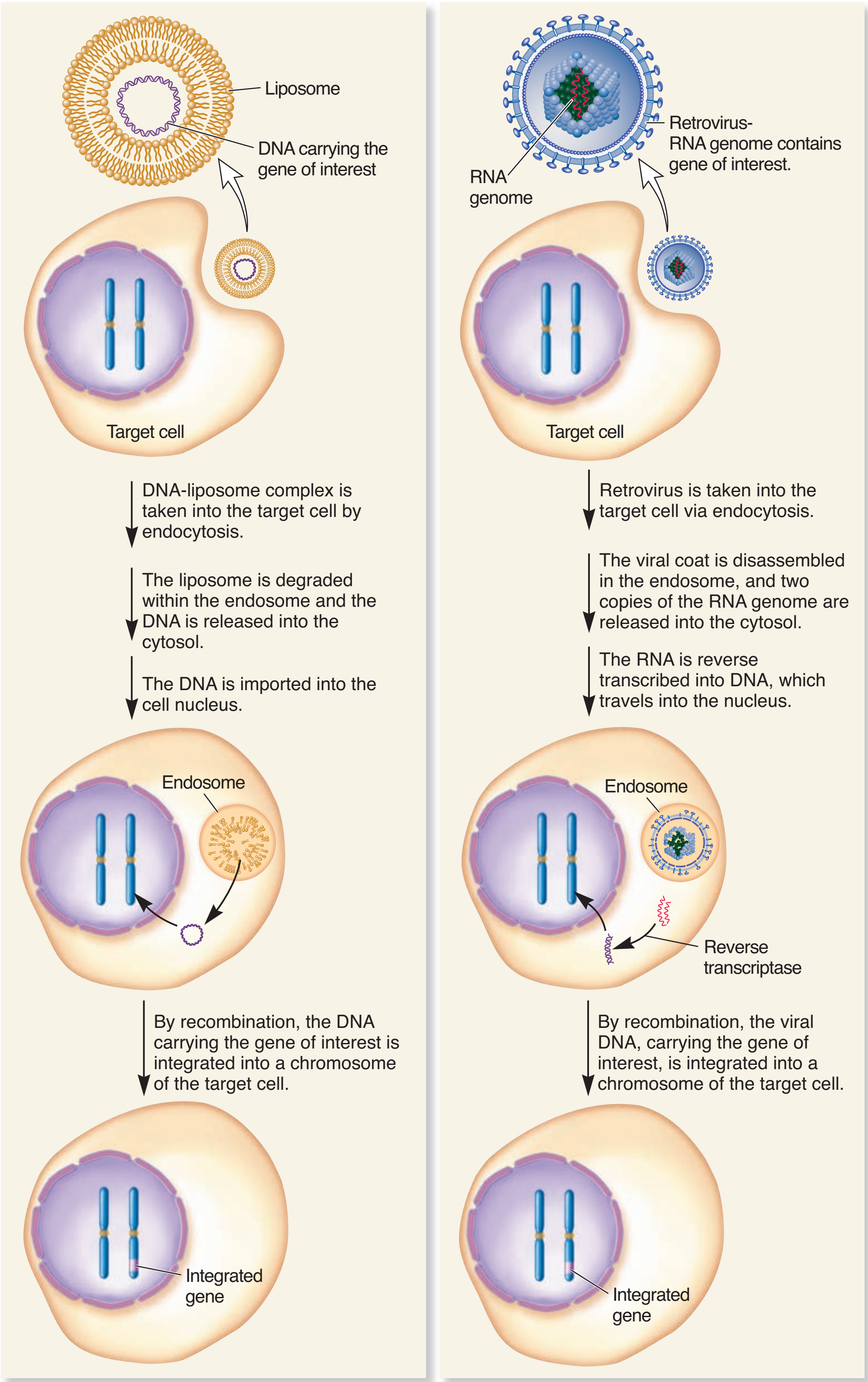
Adenosine Deaminase Deficiency Was the First Inherited Disease Treated with Gene Therapy

Adenosine deaminase (ADA) is an enzyme involved in purine metabolism. If both copies of the *ADA* gene are defective, deoxyadenosine accumulates within the cells of the individual. At high concentrations, deoxyadenosine is particularly toxic to lymphocytes in the immune system, namely, T cells and B cells. In affected individuals, the destruction of T and B cells leads to a form of severe combined immunodeficiency (SCID), in which the immune system's ability to fight infectious disease is compromised or absent. If left untreated, SCID is typically fatal at an early age (generally, 1–2 years old), because the immune system of these individuals is severely compromised and cannot fight infections.

Three approaches can be used to treat ADA deficiency. In some cases, a patient may receive a bone marrow transplant from a compatible donor. A second method is to treat SCID patients with purified ADA that is coupled to polyethylene glycol (PEG). This PEG-ADA is taken up by lymphocytes and can correct the ADA deficiency. Unfortunately, these two approaches are not always available and/or successful.

A third, more recent approach is to treat ADA patients with gene therapy. On September 14, 1990, the first human gene therapy was approved for a young girl suffering from ADA deficiency. This work was carried out by a large team of researchers including R. Michael Blaese, Kenneth Culver, W. French Anderson, and colleagues. Prior to this clinical trial, the normal gene for ADA had been cloned into a retroviral vector that can infect lymphocytes. The general aim of this therapy was to remove lymphocytes from the blood of the young girl with SCID, introduce the normal *ADA* gene into her cells, and then return them to her bloodstream.

Figure 20.18 outlines the protocol for the experimental treatment. Lymphocytes (including T cells) were removed and cultured in a laboratory. The lymphocytes were then transfected


Liposome
DNA carrying the gene of interest
Target cell
DNA-liposome complex is taken into the target cell by endocytosis.
The liposome is degraded within the endosome and the DNA is released into the cytosol.
The DNA is imported into the cell nucleus.
Endosome
By recombination, the DNA carrying the gene of interest is integrated into a chromosome of the target cell.
Integrated gene
RNA genome
Retrovirus-RNA genome contains gene of interest.
Target cell
Retrovirus is taken into the target cell via endocytosis.
The viral coat is disassembled in the endosome, and two copies of the RNA genome are released into the cytosol.
The RNA is reverse transcribed into DNA, which travels into the nucleus.
Endosome
Reverse transcriptase
By recombination, the viral DNA, carrying the gene of interest, is integrated into a chromosome of the target cell.
Integrated gene

(a) Nonviral approach **(b) Viral approach**

FIGURE 20.17 Methods of gene transfer used in gene therapy. (a) In the nonviral approach, the DNA containing the gene of interest is complexed with cationic liposomes. These complexes are taken into cells by endocytosis, in which a portion of the plasma membrane invaginates and creates an intracellular vesicle known as an endosome. After it is released from the endosome, the DNA may then integrate into the chromosomal DNA via recombination. **(b)** In the viral approach, the gene of interest is cloned into a retrovirus. When the retrovirus infects a cell, the RNA genome is reverse-transcribed into double-stranded DNA, which then integrates into the chromosome. Viruses used in gene therapy have been genetically altered so they cannot proliferate after entry into the target cell.

Concept check: What are advantages and disadvantages of these two methods?

with a nonpathogenic retrovirus that had been genetically engineered to contain the normal *ADA* gene. During the life cycle of a retrovirus, the retroviral genetic material is inserted into the host cell's DNA. Therefore, because this retrovirus contained the normal ADA gene, this gene also was inserted into the chromosomal DNA of the girl's lymphocytes. After this had occurred in the laboratory, the cells were reintroduced back into the patient. This method is called an **ex vivo approach** because the genetic manipulations occur outside the body, and the products are reintroduced into the body.

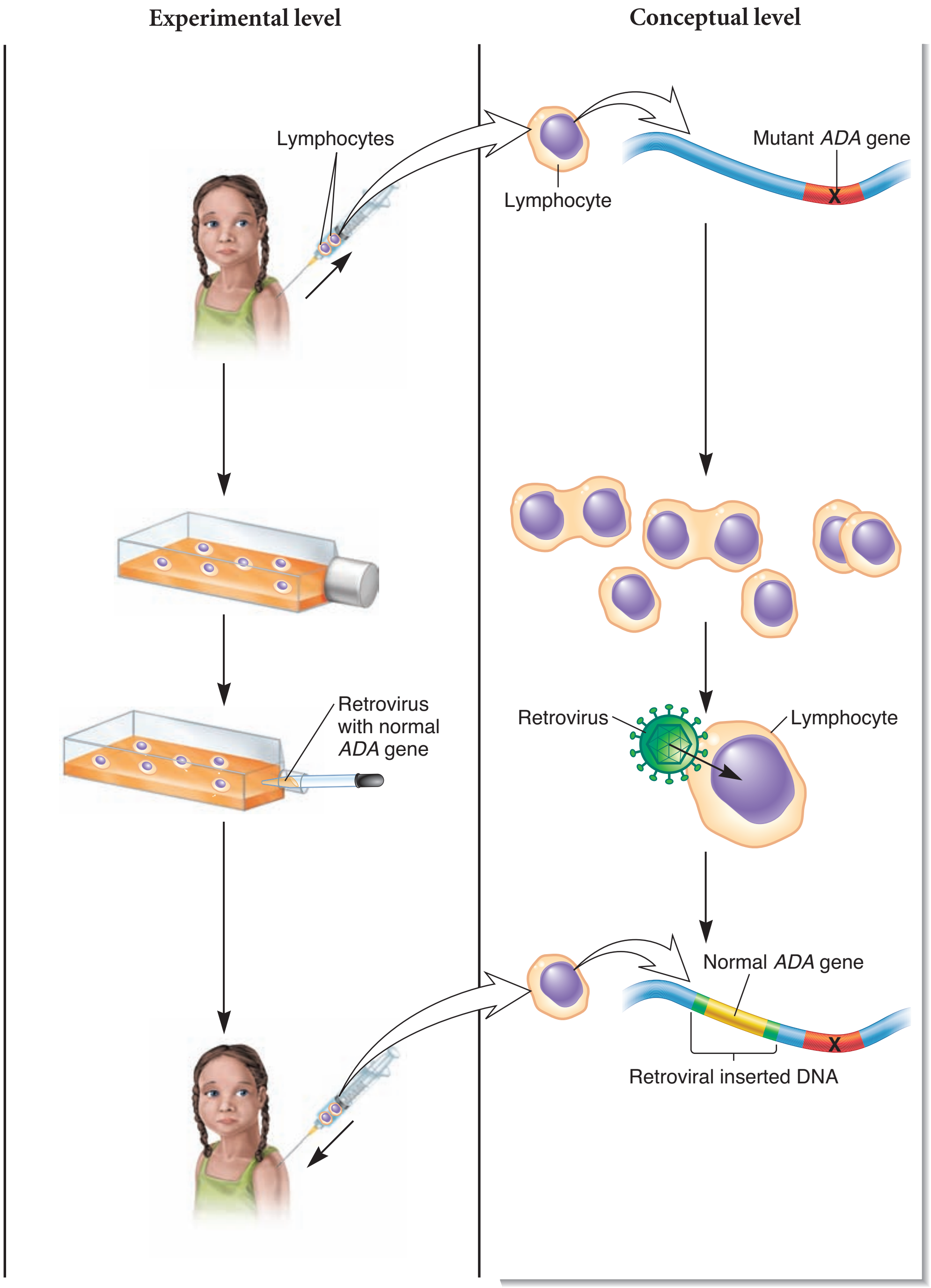
THE HYPOTHESIS

Infecting lymphocytes with a retrovirus containing the normal *ADA* gene will correct the inherited deficiency of the mutant *ADA* gene in patients with ADA deficiency.

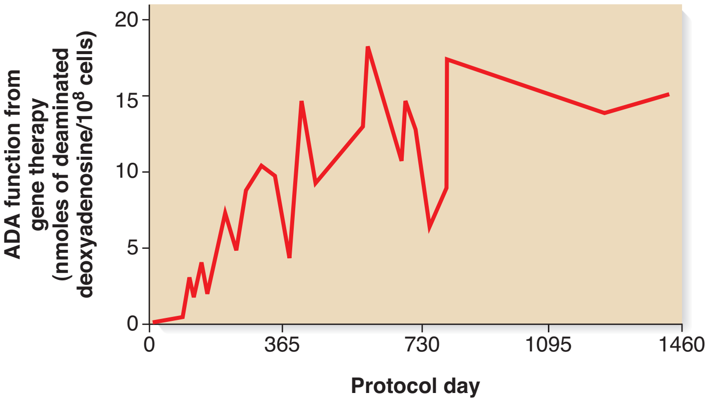
TESTING THE HYPOTHESIS — FIGURE 20.18 The first human gene therapy for adenosine deaminase deficiency by Blaese and colleagues.

Starting material: A retrovirus carrying the normal *ADA* gene.

1. Remove *ADA*-deficient lymphocytes from the patient with severe combined immunodeficiency disease (SCID).
2. Culture the cells in a laboratory.
3. Infect the cells with a nonpathogenic retrovirus that contains the normal *ADA* gene. Retroviruses insert their DNA into the host cell chromosome as part of their reproductive cycle.
4. Infuse the *ADA*-gene-corrected lymphocytes back into the SCID patient.


Experimental level
Conceptual level
Lymphocytes
Lymphocyte
Mutant ADA gene
X
Retrovirus with normal ADA gene
Retrovirus
Lymphocyte
Normal ADA gene
X
Retroviral inserted DNA

THE DATA



Data from R. M. Blaese, K. W. Culver, A. D. Miller, et al. (1995) T lymphocyte-directed gene therapy for ADA-SCID: Initial trial results after 4 years. *Science* 270:475–480.

INTERPRETING THE DATA

In this clinical trial, two patients were enrolled, and a third patient was later treated in Japan. The data show the results from one of those patients. Was the treatment a success? In at least one patient, T cells carrying the cloned gene were still detectable 4 years after they had been transferred. However, most of the circulating T cells were not found to contain the cloned gene. Because the three patients also received a low dose of PEG-ADA treatment, researchers could not determine whether or not gene transfer into T cells by itself was of significant clinical benefit.

Another form of SCID, termed SCID-X1, is inherited as an X-linked trait. SCID-X1 is characterized by a block in T-cell growth and differentiation. This block is caused by mutations in the gene encoding the γ_c cytokine receptor, which plays a key role in the recognition of signals that are needed to promote the growth, survival, and differentiation of T cells. A gene therapy trial for SCID-X1 similar to the trial shown in Figure 20.18 was initiated in 2000 in which a normal γ_c cytokine receptor gene was cloned into a retroviral vector and then introduced into SCID-X1 patients' lymphocytes. The lymphocytes were then reintroduced back into their bodies. At a 10-month follow-up, T cells expressing the normal γ_c cytokine receptor were detected in two patients. Most importantly, the T-cell counts in these two patients had risen to levels that were comparable to those in normal individuals. This clinical trial was the first clear demonstration that gene therapy can offer clinical benefit, providing in these cases what seemed to be a complete correction of the disease phenotype. However, in a French study involving 10 SCID-X1 patients, an unexpected and serious side effect occurred. Within 3 years of gene therapy treatment, 3 out of the 10 treated children developed leukemia—a form of cancer involving the proliferation of white blood cells. In these cases, the disease was caused by the integration of the retroviral vector next to a particular gene in the patients' genomes. The development of leukemia in these patients has halted many clinical trials involving gene therapy.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

20.5 REVIEWING THE KEY CONCEPTS

- Human gene therapy is the introduction of cloned genes into cells in an attempt to treat a disease. Genes may be introduced via liposomes or viruses (see Figure 20.17).
- The first human gene therapy trial was aimed at treating adenosine deaminase deficiency (see Figure 20.18).

20.5 COMPREHENSION QUESTIONS

1. A method of introducing a cloned gene into cells for gene therapy is via
   a. liposomes.
   b. retroviral vectors.
   c. T-DNA vectors.
   d. both a and b.
2. Which of the following best describes the approach that was used in the first gene therapy trial for SCID?
   a. The normal ADA gene was introduced by injecting liposomes directly into the patients' bodies.
   b. Lymphocytes were removed from a SCID patient, the normal *ADA* gene was transferred into the lymphocytes via liposomes, and then the lymphocytes were returned to the patient's body.
   c. Lymphocytes were removed from a SCID patient, the normal *ADA* gene was transferred into the lymphocytes via a retrovirus, and then the lymphocytes were returned to the patient's body.
   d. None of the above

KEY TERMS

Page 474. biotechnology, genetically modified organisms (GMOs), transgenic organism, transgene
Page 476. biological control, bioremediation, biotransformation, biodegradation
Page 478. gene replacement, gene knockout, gene addition
Page 479. chimera
Page 481. gene redundancy, gene knockin, mouse model
Page 482. molecular pharming
Page 483. reproductive cloning
Page 485. stem cells, totipotent, embryonic stem cells (ES cells), pluripotent, embryonic germ cells (EG cells), multipotent, unipotent
Page 488. T DNA, T-DNA vectors, biolistic gene transfer, microinjection, electroporation, protoplasts
Page 491. gene therapy
Page 492. liposomes
Page 494. ex vivo approach

CHAPTER SUMMARY

- Biotechnology is the application of technologies that involve the use of living organisms, or products from living organisms, for the development of products that benefit humans. Genetically modified organisms have received genetic material via recombinant DNA technology.

20.1 Uses of Microorganisms in Biotechnology

- Microorganisms have several uses in biotechnology (see Table 20.1).
- Recombinant microorganisms are used to make human hormones, such as insulin (see Table 20.2, Figure 20.1).
- Microorganisms may be used as biological control agents. In bioremediation, living organisms are used to decrease pollutants in the environment.

20.2 Genetically Modified Animals

- A transgenic organism carries a gene from another species (see Figure 20.2).
- The integration of a cloned gene into a chromosome may result in gene replacement or gene addition (see Figures 20.3, 20.4).
- Gene knockouts and knockins in mice are made to study gene function and to produce mouse models for human diseases (see Figures 20.5, 20.6).
- Transgenic livestock may produce human proteins, such as hormones, in their milk (see Table 20.3, Figure 20.7).

20.3 Reproductive Cloning and Stem Cells

- Mammalian reproductive cloning can be achieved by fusing somatic cells to oocytes with their nuclei removed (see Figures 20.8, 20.9).
- Stem cells have the ability to divide and differentiate. Stem cells may be totipotent, pluripotent, multipotent, or unipotent (see Figures 20.10–20.12).
- Stem cells have the potential to treat a variety of human diseases (see Table 20.4).

20.4 Genetically Modified Plants

- *Agrobacterium tumefaciens* transfers T DNA to plants. Researchers have used T-DNA vectors to make genetically modified plants (see Figures 20.13, 20.14).
- Researchers have made many transgenic plants that have traits that are useful to humans, including herbicide and pesticide resistance (see Table 20.5, Figures 20.15, 20.16).

20.5 Human Gene Therapy

- Human gene therapy is the introduction of cloned genes into cells in an attempt to treat a disease. Genes may be introduced via liposomes or viruses (see Figure 20.17).
- The first human gene therapy trial was aimed at treating adenosine deaminase deficiency (see Figure 20.18).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Which of the following would appropriately be described as a transgenic organism?

A. The sheep Dolly, which was produced by cloning

B. A sheep that produces human α_1-antitrypsin in its milk

C. Bt corn

D. A hybrid strain of corn produced from crossing two inbred strains of corn (The inbred strains were not transgenic.)

Answer:

A. No, Dolly was not produced using recombinant techniques. Pieces of DNA were not cut and combined in a new way.

B. Yes

C. Yes

D. No, the hybrids simply contain chromosomal genes from two different parental strains.

S2. Describe the strategy for producing human proteins in the milk of livestock.

Answer: Milk proteins are encoded by genes with promoters and regulatory sequences that direct the expression of these genes within the cells of the mammary gland. To get other proteins expressed in the mammary gland, the strategy is to link the promoter and regulatory sequences from a milk-specific gene to the coding sequence of the gene that encodes the human protein of interest. In some cases, it is also necessary to add a signal sequence to the amino-terminal end of the target protein. A signal sequence is a short polypeptide that directs the secretion of a protein from a cell. If the target protein does not already have a signal sequence, it is possible to use a signal sequence from a milk-specific gene to promote the secretion of the target protein from the mammary cells and into the milk. During this process, the signal sequence is cleaved from the secreted protein.

S3. With regard to genetically modified organisms, describe unsuccessful and successful examples.

Answer: An unsuccessful example was an oil-eating bacterium. Two very successful examples include bacteria that make human insulin and Bt varieties of agricultural crops.

Conceptual Questions

C1. What is a recombinant microorganism? Discuss examples of recombinant microbes.

C2. A conjugation-deficient strain of *A. radiobacter* is used to combat crown gall disease. Explain how this bacterium prevents the disease and the advantage of a conjugation-deficient strain.

C3. What is bioremediation? What is the difference between biotransformation and biodegradation?

C4. What is a biological control agent? Briefly describe three examples.

C5. As described in Table 20.2, several medical agents are now commercially produced by genetically engineered microorganisms. Discuss the advantages and disadvantages of making these agents this way.

C6. What is a mouse model for human disease?

C7. What is a transgenic organism? Describe three examples.

C8. What part of the *A. tumefaciens* DNA gets transferred to the genome of a plant cell during infection?

C9. Explain the difference between gene addition and gene replacement. Are the following descriptions examples of gene addition or gene replacement?

A. A mouse model to study cystic fibrosis

B. Introduction of a pesticide-resistance gene into corn using the T-DNA vector of *A. tumefaciens*

C10. As described in Chapter 6, not all inherited traits are determined by nuclear genes (i.e., genes located in the cell nucleus) that are expressed during the life of an individual. In particular, maternal effect genes and mitochondrial genes are notable exceptions. With these ideas in mind, let's consider the cloning of sheep (e.g., Dolly).

A. With regard to maternal effect genes, is the phenotype of such a cloned animal determined by the animal that donated the enucleated egg or by the animal that donated the somatic cell nucleus? Explain.

B. Does the cloned animal inherit extranuclear traits from the animal that donated the egg or from the animal that donated the somatic cell? Explain.

C. In what ways would you expect this cloned animal to be similar to or different from the animal that donated the somatic cell? Is it accurate to call such an animal a "clone" of the animal that donated the nucleus?

C11. Discuss some of the worthwhile traits that can be modified in transgenic plants.

C12. Discuss the concerns that some people have about the uses of genetically engineered organisms.

Application and Experimental Questions

E1. Recombinant bacteria can produce hormones that are normally produced in humans. Briefly describe how this is accomplished.

E2. *Bacillus thuringiensis* can make toxins that kill insects. This toxin must be applied several times during the growth season to prevent insect damage. As an alternative to repeated applications, one strategy is to apply bacteria directly to leaves. However, *B. thuringiensis* does not survive very long in the field. Other bacteria, such as *Pseudomonas syringae*, do. Propose a way to alter *P. syringae* so it could be used as an insecticide. Discuss advantages and disadvantages of this approach compared with the repeated applications of the insecticide from *B. thuringiensis*.

E3. In the experiment of Figure 20.1, why was it necessary to link the coding sequence for the A or B chains to the sequence for β-galactosidase? How were the A or B chains separated from β-galactosidase after the fusion protein was synthesized in *E. coli*?

E4. Explain how it is possible to select for homologous recombination in mice. What phenotypic marker is used to readily identify chimeric mice?

E5. To produce transgenic plants, plant tissue is exposed to *Agrobacterium tumefaciens* and then grown in media containing kanamycin, carbenicillin, and plant growth hormones. Explain the purpose behind each of these three agents. What would happen if you left out the kanamycin?

E6. List and briefly describe five methods for the introduction of cloned genes into plants.

E7. What is a gene knockout? Is an animal or plant with a gene knockout a heterozygote or homozygote? What might you conclude if a gene knockout does not have a phenotypic effect?

E8. It is common for researchers to identify genes using cloning methods described in Chapters 19 and 20. A gene can be identified according to its molecular features. For example, a segment of DNA can be identified as a gene because it contains the right combination of sequences: a promoter, exons, introns, and a terminator. Or a gene can be identified because it is transcribed into mRNA. In the study of plants and animals, it is relatively common for researchers to identify genes using molecular techniques without knowing the function of the gene. In the case of mice, the function of the gene can be investigated by making a gene knockout. A knockout that causes a phenotypic change in the mouse may provide an important clue regarding the function of a gene. For example, a gene knockout that produced an albino mouse would indicate the gene knocked out probably plays a role in pigment formation. The experimental strategy of first identifying a gene based on its molecular properties and then investigating its function by making a knockout is called reverse genetics. Explain how this approach is opposite (or "in reverse") to the conventional way that geneticists study the function of genes.

E9. According to the methods described in Figure 20.5, can homologous recombination that results in gene replacement cause the integration of both the *TK* and Neo^R genes? Explain why or why not. Describe how the *TK* gene and Neo^R gene are used in a selection scheme that favors gene replacement.

E10. What is a chimera? How are chimeras made?

E11. Evidence (see P. G. Shiels, A. J. Kind, K. H. Campbell, et al. (1999) Analysis of telomere lengths in cloned sheep. *Nature 399,* 316–17) shows that Dolly may have been "genetically older" than her actual age would have suggested. As mammals age, the chromosomes in somatic cells tend to shorten from the telomeres. Therefore, older individuals have shorter chromosomes in their somatic cells than do younger ones. When researchers analyzed the chromosomes in the somatic cells of Dolly when she was about 3 years old, the lengths of her chromosomes were consistent with a sheep that was significantly older, say, 9 or 10 years old. (Note: As described in this chapter, the sheep that donated the somatic cell that produced Dolly was 6 years old, and her mammary cells had been grown in culture for several cell doublings before a mammary cell was fused with an oocyte.)

A. Suggest an explanation why Dolly's chromosomes seemed so old.

B. Let's suppose that Dolly at age 11 gave birth to a lamb named Molly; Molly was produced naturally (by mating Dolly with a normal male). When Molly was 8 years old, a sample of somatic cells was analyzed. How old would you expect Molly's chromosomes to appear, based on the phenomenon of telomere shortening? Explain your answer.

C. Discuss how the observation of chromosome shortening, which was observed in Dolly, might affect the popularity of reproductive cloning.

E12. When transgenic organisms are made, the transgene may integrate into multiple sites within the genome. Furthermore, the integration site may influence the expression of the gene. For example, if a transgene integrates into a heterochromatic (i.e., highly condensed) region of a chromosome, the transgene may not be expressed. For these reasons, it is important for geneticists to analyze transgenic organisms with regard to the number of transgene insertions and the expression levels of the transgenes. Chapter 19 describes three methods (Southern blotting, Northern blotting, and Western blotting) that can be used to detect genes and gene products. Which of these techniques would you use to determine the number of transgenes in a transgenic plant or animal? Why is it important to know the number of transgenes? (Hint: You may want to use a transgenic animal or plant as breeding stock to produce many more transgenic animals or plants.) Which technique would you use to determine the expression levels of transgenes?

E13. What is molecular pharming? Compared with the production of proteins by bacteria, why might it be advantageous?

E14. What is reproductive cloning? Are identical twins in humans considered to be clones? With regard to agricultural species, what are some potential advantages to reproductive cloning?

E15. Researchers have identified a gene in humans that (when mutant) causes severe dwarfism and mental deficiencies. This disorder is inherited in an autosomal recessive manner, and the mutant allele is known to be a loss-of-function mutation. The same gene has been found in mice, although a mutant version of the gene has not been discovered in mice. To develop drugs and an effective therapy to treat this disorder in humans, it would be experimentally useful to have a mouse model. In other words, it would be desirable to develop a strain of mice that carry the mutant allele in the homozygous condition. Experimentally, how would you develop such a strain?

E16. Treatment of adenosine deaminase (ADA) deficiency is an example of ex vivo gene therapy. Why is this therapy called ex vivo? Can ex vivo gene therapy be used to treat all inherited diseases? Explain.

E17. Several research studies are under way that involve the use of gene therapies to inhibit the growth of cancer cells. As discussed in Chapter 23, oncogenes are mutant genes that are overexpressed and cause cancer. New gene therapies are aimed at silencing an oncogene by producing antisense RNA that recognizes the mRNA transcribed from an oncogene. Based on your understanding of antisense RNA (described in Chapters 16 and 17), explain how this strategy would prevent the growth of cancer cells.

Questions for Student Discussion/Collaboration

1. Discuss the advantages and disadvantages of gene therapy. Because a limited amount of funding is available for gene therapy research, make a priority list of the three top diseases for which you would fund research. Discuss your choices.

2. Make a list of the types of traits you would like to see altered in transgenic plants and animals. Suggest ways (i.e., what genes would you use?) to accomplish these alterations.

Answers to Comprehension Questions

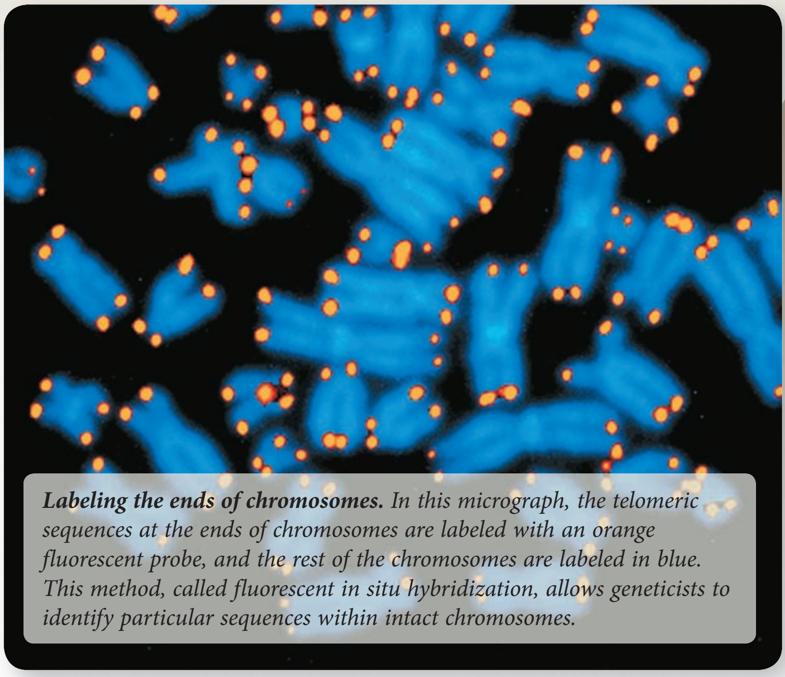
20.1: d, b, c
20.2: a, b, b
20.3: c, d
20.4: b, d
20.5: d, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

21



Labeling the ends of chromosomes. *In this micrograph, the telomeric sequences at the ends of chromosomes are labeled with an orange fluorescent probe, and the rest of the chromosomes are labeled in blue. This method, called fluorescent in situ hybridization, allows geneticists to identify particular sequences within intact chromosomes.*

CHAPTER OUTLINE

21.1 Overview of Chromosome Mapping
21.2 Cytogenetic Mapping Via Microscopy
21.3 Linkage Mapping Via Crosses
21.4 Physical Mapping Via Cloning
21.5 Genome-Sequencing Projects
21.6 Transposition

GENOMICS I: ANALYSIS OF DNA AND TRANSPOSABLE ELEMENTS

The term **genome** refers to the total genetic composition of an organism or species. For example, the nuclear genome of humans is composed of 22 different autosomes, an X chromosome, and (in males) a Y chromosome. In addition, humans have a mitochondrial genome composed of a single circular chromosome.

As genetic technology has progressed over the past few decades, researchers have gained an increasing ability to analyze the composition of genomes as a whole unit. The term **genomics** refers to the molecular analysis of the entire genome of a species. Genome analysis is a molecular dissection process applied to a complete set of chromosomes. Segments of chromosomes are analyzed in progressively smaller pieces, the locations of which are known on the intact chromosomes. This is the mapping phase of genome analysis. The mapping of the genome ultimately progresses to the determination of the complete DNA sequence for all of a species' chromosomes.

In 1995, a team of researchers headed by Craig Venter and Hamilton Smith obtained the first complete DNA sequence of the bacterial genome from *Haemophilus influenzae*. Its genome is composed of a single circular chromosome that is 1.83 million base pairs (bp) in length and contains approximately 1743 genes. In 1996, the first entire DNA sequence of a eukaryote, *Saccharomyces cerevisiae* (baker's yeast), was completed. This work was carried out by a European-led consortium of more than 100 laboratories, including some in the United States, Japan, and Canada. The overall coordinator was André Goffeau in Belgium. The yeast genome contains 16 linear chromosomes, which have a combined length of about 12.1 million bp and contain approximately 6300 genes. Since that time, genome sequences from many prokaryotes and eukaryotes have been completed.

In this chapter, we will focus on methods aimed at elucidating the organization of the sequences within a species' genome. This process may begin with the mapping of regions along a species' chromosomes. The process is finished when the complete DNA sequence has been determined. We will consider three mapping strategies—cytogenetic, linkage, and physical mapping—and the approaches used to carry them out. Then we explore genome-sequencing projects—research endeavors that have the ultimate goal of determining the sequence of DNA bases of the entire genome of a given species. We will examine the methods, goals, and results of these large undertakings, which include the Human Genome Project.

Once a genome sequence is known, researchers can examine, at the level of many genes, how the components of a genome interact to produce the traits of an organism. This approach is called **functional genomics.** A long-term goal of researchers is to determine the roles of all cellular proteins, as well as the interactions that these proteins experience, to produce the characteristics of particular cell types and the traits of complete organisms. This is the research area known as **proteomics.** In this chapter, we explore genomics at the level of DNA segments and sequences. In Chapter 22, we will consider the exciting prospects that functional genomics and proteomics have to offer regarding our future understanding of genetics.

21.1 OVERVIEW OF CHROMOSOME MAPPING

In genetics, the term **mapping** refers to the experimental process of determining the relative locations of genes or other segments of DNA along individual chromosomes. Researchers may follow three general approaches to mapping a chromosome: cytogenetic, linkage, and physical mapping strategies. Before we discuss these approaches in detail, let's compare them. **Cytogenetic mapping** (also called cytological mapping) relies on the localization of gene sequences within chromosomes that are viewed microscopically. When stained, each chromosome of a given species has a characteristic banding pattern, and genes are mapped cytogenetically relative to a band location. By comparison, in Chapter 7, we considered how genetic crosses are conducted to map the relative locations of genes along a chromosome. Such genetic studies, which are called **linkage mapping,** use the frequency of genetic recombination between different genes to determine their relative spacing and order along a chromosome. In eukaryotes, linkage mapping involves crosses among organisms that are heterozygous for two or more genes. The number of recombinant offspring provides a relative measure of the distance between genes, which is computed in map units (or centiMorgans, cM). Finally, a third approach is **physical mapping,** in which DNA-cloning techniques are used to determine the location and distance between genes and other DNA regions. In a physical map, the distances are computed as the number of base pairs between genes.

A **genetic map,** or **chromosome map,** is a chart that describes the relative locations of genes or other DNA segments along a chromosome. The term **locus** (plural, **loci**) refers to the site within a genetic map where a specific gene or other DNA segment is found. Figure 21.1 compares genetic maps that show the loci for two X-linked genes, *sc* (scute, a gene affecting bristle morphology) and *w* (a gene affecting eye color), in *Drosophila melanogaster.* In the cytogenetic map (top), the *sc* gene is located at band 1A8, and the *w* gene is located at band 3B6. In the linkage map, genetic crosses indicate that the two genes are approximately 1.5 map units (mu) apart. The physical map shows that the two genes are approximately 2.4 × 10^6 bp apart along the X chromosome. Correlations between cytogenetic, linkage, and physical maps often vary from species to species and from one region of the chromosome to another. For example, a distance of 1 mu may correspond to 1 to 2 million bp in one region of the chromosome, but other regions may recombine at a much lower rate, so a distance of 1 mu may be a much longer physical segment of DNA.

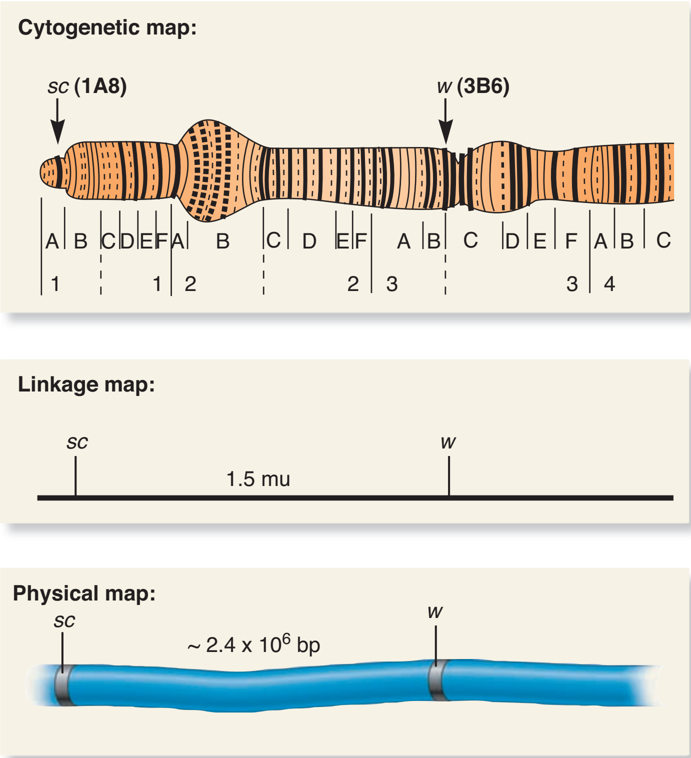


FIGURE 21.1 A comparison of cytogenetic, linkage, and physical maps. Each of these maps shows the distance between the *sc* and *w* genes along the X chromosome in *Drosophila melanogaster*. The cytogenetic map is that of the polytene chromosome.

Concept check: What is a genetic map?

21.1 REVIEWING THE KEY CONCEPTS

- The term *mapping* refers to the process of determining the relative locations of genes or other DNA segments along a chromosome, thereby producing a genetic or chromosome map. The process may involve cytogenetic, linkage, or physical mapping techniques (see Figure 21.1).

21.1 COMPREHENSION QUESTIONS

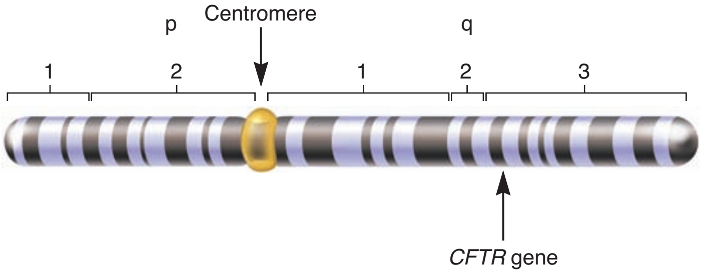
1. What type of mapping relies on microscopy?
   a. Cytogenetic mapping
   b. Linkage mapping
   c. Physical mapping
   d. All of the above

21.2 CYTOGENETIC MAPPING VIA MICROSCOPY

As we have seen, one way to determine the relative locations of genes along a chromosome is via cytogenetic mapping. This approach relies on the localization of gene sequences within chromosomes that are viewed microscopically. Cytogenetic mapping is commonly used in eukaryotes, which have very large chromosomes relative to those of bacteria. Microscopically, eukaryotic chromosomes can be distinguished from one another by their size, centromeric location, and banding patterns (refer to Chapter 8, Figure 8.1). By treating chromosomal preparations with particular dyes, a discrete banding pattern is obtained for each chromosome. Cytogeneticists use this banding pattern as a way to describe specific regions along a chromosome. In this section, we will explore techniques that are aimed at producing cytogenetic maps.

A Goal of Cytogenetic Mapping Is to Determine the Location of a Gene Along an Intact Chromosome

Cytogenetic mapping attempts to determine the locations of particular genes relative to a banding pattern of a chromosome. For example, the human gene that encodes the cystic fibrosis transmembrane regulator (CFTR) is located on chromosome 7, at a specific site in the q3 region. The CFTR protein is defective in people with cystic fibrosis.



Cytogenetic mapping may be used as a first step in the localization of genes in plants and animals. However, because it relies on light microscopy, cytogenetic analysis has a fairly crude limit of resolution. In most species, cytogenetic mapping is accurate only within limits of approximately 5 million bp along a chromosome. In species that have large polytene chromosomes, such as *Drosophila*, the resolution is much better. A common strategy used by geneticists is to roughly locate a gene by cytogenetic analysis and then determine its location more precisely by the physical mapping methods described later.

In situ Hybridization Can Localize Genes Along Particular Chromosomes

The technique of **in situ hybridization** is widely used to cytogenetically map the locations of genes or other DNA sequences along large eukaryotic chromosomes. The term *in situ* (from the Latin for "in place") indicates that the procedure is conducted on chromosomes that are being held in place—adhered to a surface.

To map a gene via in situ hybridization, researchers use a probe to detect the location of the gene within a set of chromosomes. If the gene of interest has been cloned previously, as described in Chapter 19, the DNA of the cloned gene can be used as a probe. Because a DNA strand from a cloned gene, which is a very small piece of DNA relative to a chromosome, hybridizes only to its complementary sequence on a particular chromosome, this technique provides the ability to localize the gene of interest. For example, let's consider the gene designated *w* that causes the white-eye phenotype in *Drosophila* when it carries a loss-of-function mutation. This gene has already been cloned. If a single-stranded piece of this cloned DNA is mixed with *Drosophila* chromosomes in which the DNA has been denatured, it will bind only to the X chromosome at the location corresponding to the site of the eye color gene.

The most common method of in situ hybridization uses fluorescently labeled DNA probes and is referred to as **fluorescence in situ hybridization (FISH)**. Figure 21.2 describes the steps of the FISH procedure. The cells are prepared using a technique that keeps the chromosomes intact. The cells are then treated with agents that cause them to swell, and their contents are fixed to the slide. The chromosomal DNA is then denatured, and a DNA probe is added. For example, the added DNA probe might be single-stranded DNA that is complementary to a specific gene. In this case, the goal of a FISH experiment is to determine the location of the gene within a set of chromosomes. The probe binds to a site in the chromosomes where the gene is located because the probe and chromosomal gene line up and their complementary sequences hydrogen bond with each other. To detect where the probe has bound to a chromosome, the probe is subsequently tagged with a fluorescent molecule. This is usually accomplished by first incorporating biotin-labeled nucleotides into the probe. Biotin, a small, nonfluorescent molecule, has a very high affinity for a protein called avidin. Fluorescently labeled avidin is added, which binds tightly to the biotin, thereby labeling the probe as well.

How is the fluorescently labeled probe detected? A fluorescent molecule absorbs light at a particular wavelength and then emits light at a longer wavelength. To detect the light emitted by a fluorescently labeled probe, a fluorescence microscope is used. Such a microscope contains filters that allow the passage of light only within a defined wavelength range. The sample is illuminated at the wavelength of light that is absorbed by the fluorescent molecule. The fluorescent molecule then emits light at a longer wavelength. The fluorescence microscope has a filter that allows the transmission of the emitted light. Because only the emitted light is viewed, the background of the sample is dark, and the fluorescence is seen as a brightly glowing color on a dark background. For most FISH experiments, chromosomes are generally counterstained by a fluorescent dye that is specific for DNA. A commonly used dye is DAPI (4′,6-diamidino-2-phenylindol), which is excited by UV light. This provides all of the DNA with a blue background. The results of a FISH experiment are then compared with a sample of chromosomes that have been stained with Giemsa to produce banding, so the location of a probe can be mapped relative to the banding pattern.

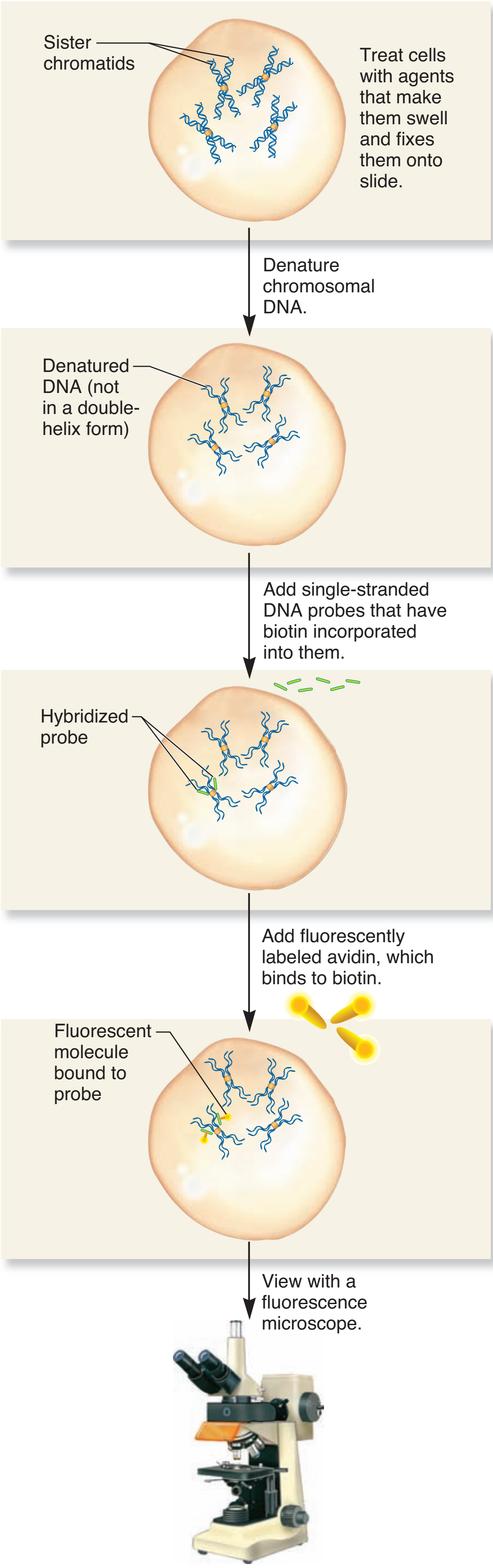

Sister chromatids
Treat cells with agents that make them swell and fixes them onto slide.
Denature chromosomal DNA.
Denatured DNA (not in a double-helix form)
Add single-stranded DNA probes that have biotin incorporated into them.
Hybridized probe
Add fluorescently labeled avidin, which binds to biotin.
Fluorescent molecule bound to probe
View with a fluorescence microscope.

FIGURE 21.2 **The technique of fluorescence in situ hybridization (FISH).** The probe hybridizes to the denatured chromosomal DNA only at specific, complementary sites in the genome. Note that the chromosomes are highly condensed metaphase chromosomes that have already replicated. They are sister chromatids. Therefore, each X-shaped chromosome actually contains two copies of a particular gene. Because the sister chromatids are identical, a probe that recognizes one sister chromatid also binds to the other.

Concept check: Why does the probe bind to a specific site on a chromosome?

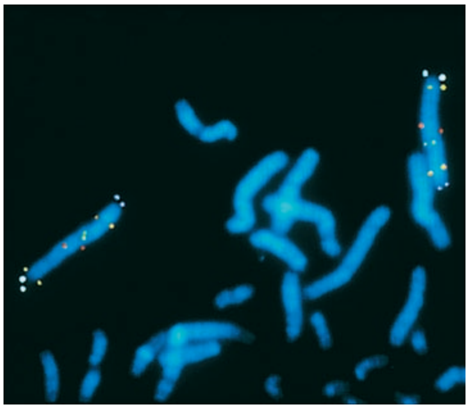


FIGURE 21.3 **The results of a fluorescence in situ hybridization experiment.** In this experiment, six different probes were used to locate six different sites along human chromosome 5. The colors are due to computer imaging of fluorescence emission; they are not the actual colors of the fluorescent labels. Two spots are usually seen at each site because the probe binds to both sister chromatids.

Figure 21.3 illustrates the results of an experiment involving six different DNA probes. The six probes were strands of DNA corresponding to six different DNA segments of human chromosome 5. In this experiment, each probe was labeled with a different fluorescent molecule. This enabled researchers to distinguish the probes when they became bound to their corresponding locations on chromosome 5. In this experiment, computer-imaging methods were used to assign each fluorescently labeled probe a different color. This method is called **chromosome painting.** In this way, FISH discerns the sites along chromosome 5 corresponding to the six different probes. In a visual, colorful way, FISH was used here to determine the order and relative distances between several specific sites along a single chromosome. FISH is commonly used in genetics and cell biology research, and its use has become more widespread in clinical applications. For example, clinicians may use FISH to detect changes in chromosome structure such as deletions, duplications, and translocations, which may occur in patients with genetic disorders.

21.2 REVIEWING THE KEY CONCEPTS

- Cytogenetic mapping attempts to determine the locations of particular genes relative to a banding pattern of a chromosome.

- Fluorescence in situ hybridization (FISH) is a commonly used method to map genes and other segments along a chromosome (see Figures 21.2, 21.3).

21.2 COMPREHENSION QUESTIONS

1. The technique of fluorescence in situ hybridization involves the use of a ______________ that hybridizes to a ______________.
   a. radiolabeled probe, band on a gel
   b. radiolabeled probe, specific site on an intact chromosome
   c. fluorescent probe, band on a gel
   d. fluorescent probe, specific site on an intact chromosome

21.3 LINKAGE MAPPING VIA CROSSES

Let's now turn to linkage mapping, which relies on the frequency of recombinant offspring to determine the distance between sites located along the same chromosome. We considered linkage mapping methods in Chapter 6, where testcrosses involving allelic differences between genes were conducted to map the relative locations of those genes along a chromosome. In this section, we will focus on the mapping of molecular markers.

The Relative Locations of Molecular Markers Can Be Determined by Linkage Mapping

As an alternative to relying on allelic differences between genes, geneticists have realized that regions of DNA that do not encode genes can be used as markers along a chromosome. A **molecular marker** is a segment of DNA found at a specific site along a chromosome and has properties that enable it to be uniquely recognized using molecular tools such as polymerase chain reaction (PCR) and gel electrophoresis. As with alleles, the molecular markers may be **polymorphic;** that is, within a population, they may vary from individual to individual. Therefore, the distances between linked molecular markers can be determined from the outcomes of crosses. Using molecular techniques, researchers have found it easier to identify many molecular markers within a given species' genome rather than identify many allelic differences among individuals. For this reason, geneticists have increasingly turned to molecular markers as points of reference along genetic maps. As described in Table 21.1, many different kinds of molecular markers are used by geneticists.

TABLE 21.1
Common Types of Molecular Markers

Marker	Description
Restriction fragment length polymorphism (RFLP)	A site in a genome where the distance between two restriction sites varies among different individuals. These sites are identified by restriction enzyme digestion of chromosomal DNA and the use of Southern blotting.
Amplified restriction fragment length polymorphism (AFLP)	The same as an RFLP except that the fragment is amplified via PCR instead of isolating the chromosomal DNA
Minisatellite	A site in the genome that contains many repeat sequences. The total length is usually in the size range of several hundred to a few thousand base pairs. Because their total lengths are usually polymorphic, they were once used in DNA fingerprinting but have been largely superseded by microsatellites (see Chapter 25).
Microsatellite	A site in the genome that contains many short sequences that are repeated many times in a row. The total length is usually in the size range of 50–200 bp, and their lengths may be polymorphic within a population. They are isolated via PCR.
Single-nucleotide polymorphism (SNP)	A site in the genome where a single nucleotide is polymorphic among different individuals. These sites occur commonly in all genomes, and they are gaining greater use in the mapping of disease-causing alleles and in the mapping of genes that contribute to quantitative traits that are valuable in agriculture (see Chapter 26).
Sequence-tagged site (STS)	This is a general term to describe any molecular marker that is found at a unique site in the genome and is amplified by PCR. AFLPs, microsatellites, and SNPs can provide sequence-tagged sites within a genome.

Researchers have constructed detailed genetic maps in which a series of many molecular markers have been identified along each chromosome of certain species. These species include humans, model organisms, and agricultural species. Why are molecular markers useful? One key reason is that molecular markers can be used to determine the approximate location of an unknown gene that causes a human disease. Clinical geneticists sometimes follow the transmission patterns of polymorphic molecular markers in family pedigrees to identify genes that cause human disease when they are mutant. The discovery of a particular marker in those who have the disease can indicate that the marker is close to the disease-causing allele (see Figure 23.7). This may help researchers identify the gene by cloning methods, such as chromosome walking, which we will examine later in this chapter.

In addition, molecular markers may help researchers identify the locations of genes involved in quantitative traits, such as fruit yield and meat weight, that are valuable in agriculture. The use of molecular markers to identify such genes is described in Chapter 26 (see Figure 26.5). Genetic maps with a large number of markers can also be used by evolutionary biologists to determine patterns of genetic variation within a species and the evolutionary relatedness of different species.

All of the markers described in Table 21.1 have been used in linkage-mapping studies. To exemplify their use, we will first consider **restriction fragment length polymorphisms (RFLPs),** but similar mapping strategies can be followed for other types of markers. As discussed in Chapter 19, restriction enzymes recognize specific DNA sequences and cleave the DNA at those sequences. Along a very long chromosome, a particular restriction enzyme recognizes many sites. For example, a commonly used restriction enzyme, *Eco*RI, recognizes 5′-GAATTC-3′. Simply by chance, this six-nucleotide sequence is expected to occur (on average) every 4^{-6} bases, or once in every 4096 bases. A chromosome composed of millions of nucleotides contains many *Eco*RI sites, which are randomly distributed along the chromosome. Therefore, *Eco*RI digests a chromosome into many smaller pieces of various lengths.

When comparing different individuals, the digestion of chromosomal DNA by a given restriction enzyme may produce certain fragments that differ in their length, even though such fragments are found at the same chromosomal locations. In this case, geneticists would say there is polymorphism in the population with regard to the length of a particular DNA fragment. This variation can arise for different reasons. Among a population of individuals, genetic changes such as small deletions and duplications may subtract or add segments of DNA to a particular region of the chromosome. This tends to occur more frequently if a region contains a repetitive sequence. Alternatively, a mutation may change the DNA sequence in a way that alters a recognition site for a particular restriction enzyme. Such a mutation changes the sizes of certain DNA fragments that result from the digestion of the chromosomal DNA with that enzyme.

As an example, **Figure 21.4** considers the application of RFLP analysis to a short chromosomal region among three diploid individuals. Due to slight differences in their DNA sequences, these individuals vary with regard to the existence of an *Eco*RI site shown in red. In certain individuals, a mutation changed the DNA sequence at this site so it is no longer GAATTC and is not recognized by *Eco*RI. In a diploid species, each individual has two copies of this region. In individual 1, both of these *Eco*RI sites are present, but in individual 2, both are absent. By comparison, individual 3 is a heterozygote. In one of the chromosomes, the site is present, but in the other, it is absent. In this experiment, the DNA was isolated and then cut by *Eco*RI. The DNA from individual 1 is cut at six sites by *Eco*RI, whereas individual 2 has only five locations where the DNA is cut because an *Eco*RI site is missing.

When subjected to gel electrophoresis, the change in the DNA sequence at a single *Eco*RI restriction site produces a variation in the sizes of homologous DNA fragments. The term *polymorphism*, meaning many forms, refers to the idea that the individuals within a population differ with regard to these particular DNA fragments. As noted in Figure 21.4, not all restriction sites are polymorphic. The three individuals share many DNA fragments that are identical in size. When a DNA segment is identical among all members of a population, it is said to be **monomorphic** (meaning one form). As a practical rule of thumb, a DNA segment is considered monomorphic when over 99% of the individuals in the population have identical sequences at that segment.

In an actual analysis of RFLP data, researchers conduct crosses or examine pedigrees and simultaneously analyze many RFLPs of offspring to determine the likelihood of linkage between

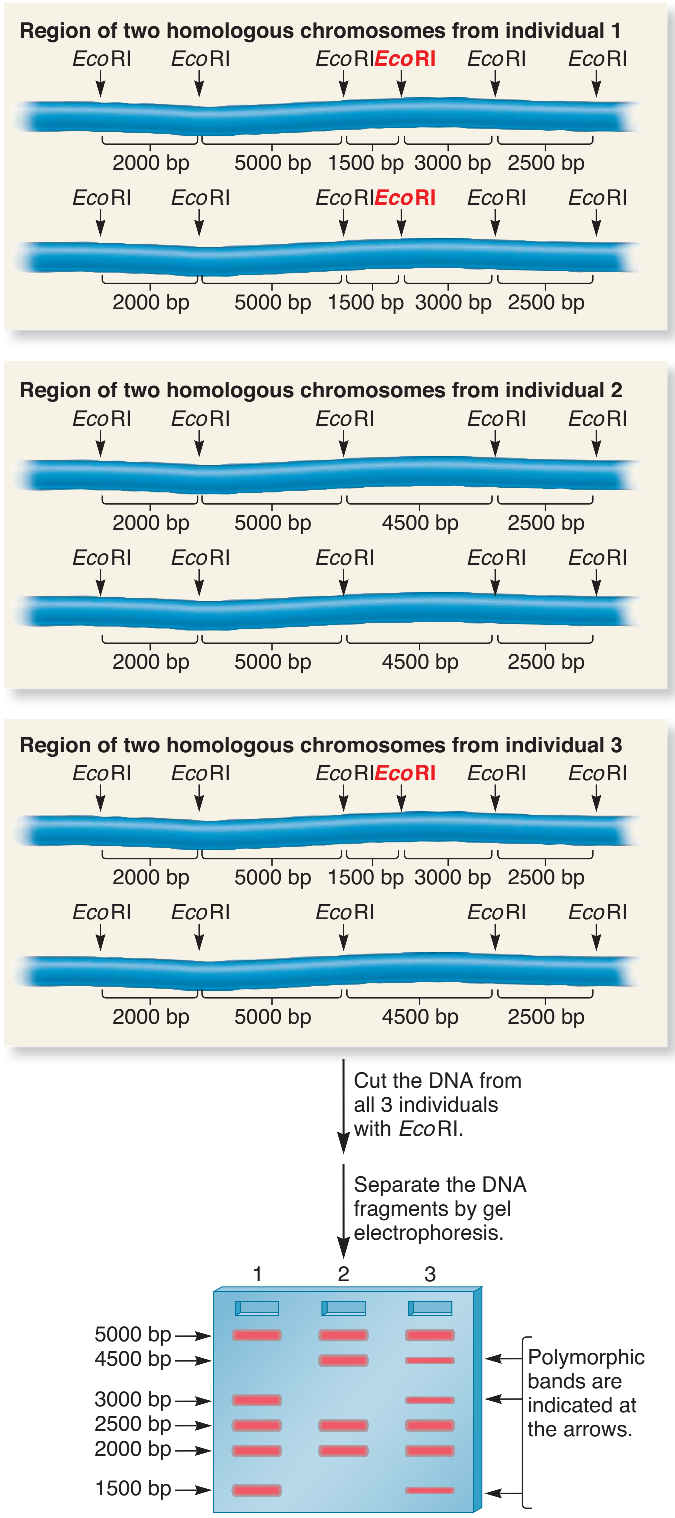

Region of two homologous chromosomes from individual 1
EcoRI
EcoRI
EcoRI
EcoRI
EcoRI
EcoRI
2000 bp
5000 bp
1500 bp
3000 bp
2500 bp
Region of two homologous chromosomes from individual 2
4500 bp
Region of two homologous chromosomes from individual 3
Cut the DNA from all 3 individuals with EcoRI.
Separate the DNA fragments by gel electrophoresis.
1
2
3
5000 bp
4500 bp
3000 bp
2500 bp
2000 bp
1500 bp
Polymorphic bands are indicated at the arrows.

FIGURE 21.4 An RFLP analysis of chromosomal DNA from three different individuals. The *Eco*RI site shown in red is sometimes missing because a mutation has changed the DNA sequence at this site.

Concept check: What is an RFLP?

RFLPs. A statistical test is used called the **lod** (logarithm of the odds) **score method.** This method was devised by Newton Morton in 1955. Although the theoretical basis of this approach is beyond the scope of this textbook, computer programs analyze pooled data from a large number of crosses or pedigrees involving many RFLPs. The programs determine the probability, or odds, that any two markers exhibit a certain degree of linkage (i.e., are within a particular number of map units apart) and also the probability that the data would have been obtained if the two markers were unlinked. The lod score is then calculated as the ratio of these values:

$$\text{Lod score} = \log_{10} \frac{\text{Probability of a certain degree of linkage}}{\text{Probability of independent assortment}}$$

For example, a lod score of 3 reflects a $\log_{10}$ of 1000. This means the probability is 1000-fold greater that the two markers are linked rather than assort independently. Traditionally, geneticists accept that two markers are linked if the lod score is 3 or greater.

Linkage Mapping Commonly Uses Molecular Markers Called Microsatellites

To make a highly refined map of a genome, many different polymorphic sites must be identified and their transmission followed from parent to offspring over many generations. RFLPs were among the first molecular markers studied by geneticists. More recently, other molecular markers have been used more commonly because they are easier to generate via PCR. As an example, let's consider **microsatellites,** also called simple sequence repeats (SSRs) or short tandem repeats (STRs), which are short repetitive sequences that are abundantly interspersed throughout a species' genome and vary in length among different individuals. Microsatellites usually contain di-, tri-, tetra-, or pentanucleotide sequences that are repeated many times in a row. For example, the most common microsatellite found in humans is a dinucleotide sequence $(CA)_n$, where n may range from 5 to more than 50. In other words, this dinucleotide sequence can be tandemly repeated 5 to 50 or more times. The $(CA)_n$ microsatellite is found, on average, about every 10,000 bases in the human genome. Researchers have identified thousands of different DNA segments that contain $(CA)_n$ microsatellites, located at many distinct sites within the human genome. Using primers complementary to the unique DNA sequences that flank a specific $(CA)_n$ region, a particular microsatellite can be amplified by PCR. In other words, the PCR primers copy only a particular microsatellite, but not the thousands of others that are interspersed throughout the genome (Figure 21.5).

When a pair of PCR primers copies a single site within a set of chromosomes, the amplified region is called a **sequence-tagged site (STS)**. When DNA is collected from a haploid cell, an STS produces only a single band on a gel. In a diploid species, an individual has two copies of a given STS. When an STS contains a microsatellite, the two PCR products may be identical and result in a single band on a gel if the region is the same length in both copies (i.e., if the individual is homozygous for the microsatellite). However, if an individual has two copies that differ in the number of repeats in the microsatellite sequence (i.e., if the

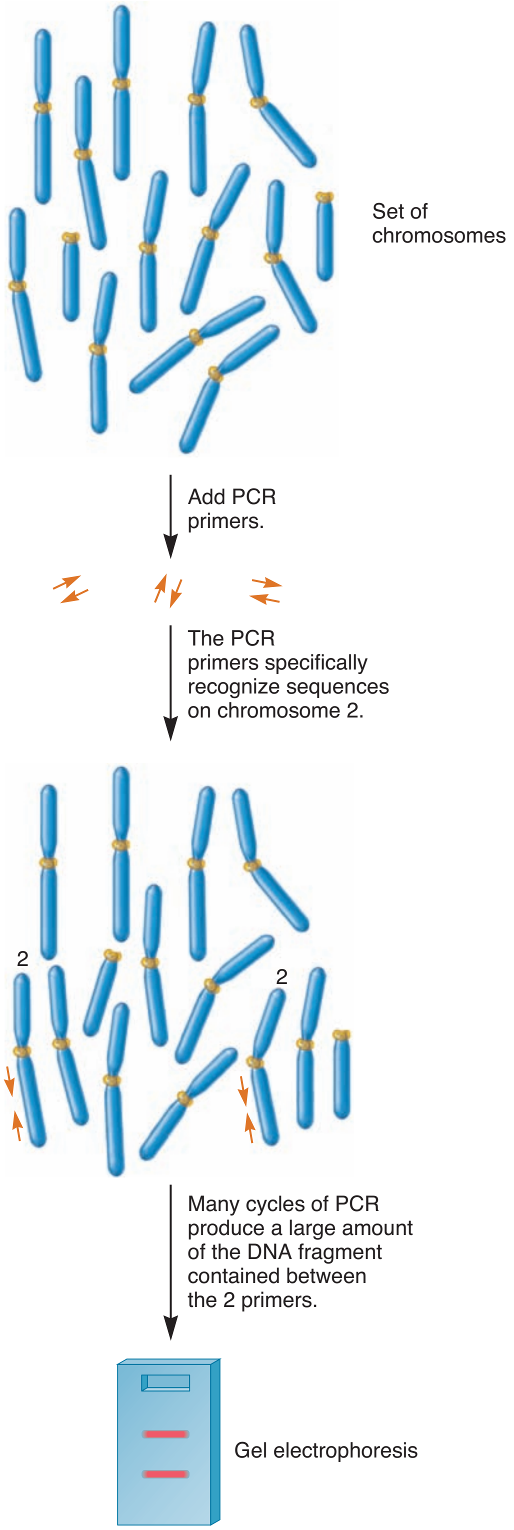

Set of chromosomes
Add PCR primers.
The PCR primers specifically recognize sequences on chromosome 2.
2
2
Many cycles of PCR produce a large amount of the DNA fragment contained between the 2 primers.
Gel electrophoresis

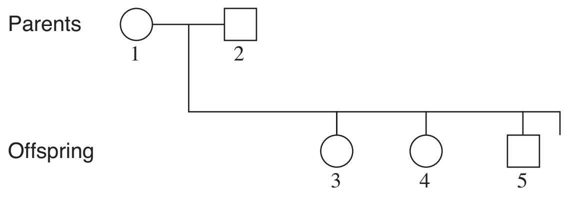
FIGURE 21.5 **Identifying a microsatellite using PCR primers.**

Concept check: What causes microsatellites to be polymorphic?

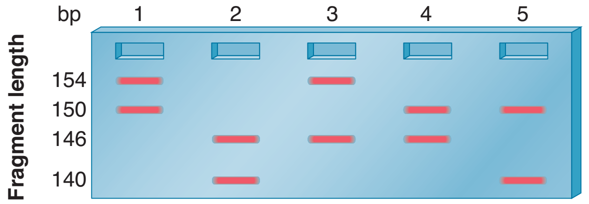
individual is heterozygous for the microsatellite), the two PCR products obtained will be different in length (as in Figure 21.5). The two bands observed in this figure differ in their masses because the same site on the two homologous chromosomes contains different numbers of repeat sequences.

Like RFLPs, microsatellites that have length polymorphisms allow researchers to follow their transmission from parent to offspring. PCR amplification of particular microsatellites provides an important strategy in the genetic analysis of human pedigrees, as shown in **Figure 21.6**. Prior to this analysis, a unique segment of DNA containing a microsatellite had been identified. Using PCR primers complementary to this microsatellite's unique flanking segments, two parents and their three offspring were tested for the inheritance of this microsatellite. A small sample of cells was obtained from each individual and subjected to PCR amplification, as described earlier in Figure 21.5. The amplified PCR products were then analyzed by high-resolution gel electrophoresis, which can detect small differences in the lengths of DNA fragments. The mother's PCR products were 154 and 150 bp in length; the father's were 146 and 140 bp. Their first offspring inherited the 154-bp product from the mother and the 146-bp one from the father, the second inherited the 150 bp from the mother and the 146 bp from the father, and the third inherited the 150 bp from the mother and the 140 bp from the father. As shown in the figure, the transmission of polymorphic microsatellites is relatively easy to follow from generation to generation.

The simple pedigree analysis shown in Figure 21.6 illustrates the general method used to follow the transmission of a single microsatellite that is polymorphic in length. In linkage studies, the goal is to follow the transmission of many different microsatellites to determine those that are linked along the same chromosome versus those that are not. Those that are not linked independently assort from generation to generation. Those that are linked tend to be transmitted together to the same offspring. In a large pedigree, it is possible to identify cases in which linked microsatellites have segregated due to crossing over. The frequency of crossing over provides a measure of the map distance, in this case between different microsatellites. This approach can help researchers obtain a finely detailed linkage map of the human chromosomes.



(a) Pedigree



(b) Electrophoretic gel of PCR products for a polymorphic microsatellite found in the family in (a).

FIGURE 21.6 **Inheritance pattern of a polymorphic microsatellite in a human pedigree.** This microsatellite is autosomal.

In addition, pedigree analysis involving STSs, such as polymorphic microsatellites, can help researchers identify the location of disease-causing alleles. The assumption behind this approach is that a disease-causing allele had its origin in a single individual known as a **founder,** who lived many generations ago. Since that time, the allele has spread throughout portions of the human population. In the case of recessive disorders, descendants of the founder who are affected with the disease have inherited two copies of the mutant allele. A second assumption is that the founder is likely to have had a polymorphic molecular marker that lies somewhere near the mutant allele. This is a reasonable assumption, because all people carry many polymorphic markers throughout their genomes. If a polymorphic marker lies very close to the disease-causing allele, it is unlikely that a crossover will occur in the intervening region. Therefore, such a polymorphic marker may be linked to the disease-causing allele for many generations. For this reason, an association between a particular polymorphic marker and a disease-causing allele may help to predict whether a person is heterozygous for a recessive disease-causing allele. By following the transmission of many polymorphic markers within large family pedigrees, it may be possible to determine that particular markers are found in people who carry specific disease-causing alleles. (An example is shown in Chapter 23, Figure 23.7.) After the identification of a closely linked marker, a disease-causing allele can be identified using a technique called chromosome walking, which is described later in this chapter.

21.3 REVIEWING THE KEY CONCEPTS

- Linkage mapping often relies on molecular markers to map the locations of genes along chromosomes (see Table 21.1).
- Restriction fragment length polymorphisms (RFLPs) and microsatellites are used as molecular markers (see Figures 21.4–21.6).

21.3 COMPREHENSION QUESTIONS

1. A molecular marker is a ______________ found at a specific site on a chromosome that has properties that allow it to be ____________.
    a. colored dye, visualized via microscopy
    b. colored dye, visualized on a gel
    c. segment of DNA, uniquely identified using molecule tools
    d. segment of DNA, visualized via microscopy
2. Which of the following is an example of a molecular marker?
    a. RFLP
    b. Microsatellite
    c. Single-nucleotide polymorphism
    d. All of the above
3. To map the distance between molecular markers via crosses, the markers must be
    a. polymorphic.
    b. monomorphic.
    c. fluorescently labeled.
    d. on different chromosomes.

21.4 PHYSICAL MAPPING VIA CLONING

We now turn our attention to methods aimed at establishing a physical map of a species' genome. Physical mapping typically involves the cloning of many pieces of chromosomal DNA. The cloned DNA fragments are then characterized by size (i.e., their length in base pairs), as well as the genes they contain and their relative locations along a chromosome.

As mentioned in Chapter 12, eukaryotic genomes are very large; the *Drosophila* genome is roughly 175 million bp long, and the human genome is approximately 3 billion bp in length. When making a physical map of a genome, researchers must characterize many DNA clones that contain much smaller pieces of the genome. In this section, we examine the general strategies used in creating a physical map of a species' genome. We will also consider how physical mapping information can be used to clone genes.

A Physical Map of a Chromosome Is Constructed by Creating a Contiguous Series of Clones That Span a Chromosome

As discussed in Chapter 19, a DNA library contains a collection of recombinant vectors in which each vector contains a particular fragment of chromosomal DNA. In physical mapping studies, the goal is to determine the relative locations of the cloned chromosomal fragments from a DNA library, as they would occur in an intact chromosome. In other words, the members of the library must be organized according to their actual locations along a chromosome. To obtain a complete physical map of a chromosome, researchers need a series of clones that contain contiguous, overlapping pieces of chromosomal DNA. Such a collection of clones is known as a **contig** (**Figure 21.7**). A contig represents a physical map of a chromosome. As discussed later, cloning vectors known as BACs and cosmids are commonly used in the construction of a contig.

Different experimental strategies can be used to align the members of a contig. The general approach is to identify clones that contain overlapping regions. Historically, Southern blotting, which is described in Chapter 19, was used to determine if two different clones contain an overlapping region. In the example shown in Figure 21.7, the DNA from clone 1 could be labeled in vitro and then used as a probe in a Southern blot of the other clones shown in this figure. Clone 1 would hybridize to clone 2, because they share identical DNA sequences in the overlapping region. Similarly, clone 2 could be used as a probe to show that it hybridizes to clone 1 and clone 3. By conducting Southern blots between many combinations of clones, researchers can determine which clones have common overlapping regions and thereby order them as they occur along the chromosome.

Alternatively, other methods for ordering the members of a contig involve the use of molecular markers. For example, if many STSs have already been identified along a chromosome via linkage or cytogenetic analysis, the STSs can be used as probes to order the members of a contig. Another approach is to subject the members of a library to a restriction enzyme to produce a restriction digestion pattern for each member. The patterns of fragments are then analyzed by computer programs, which can identify regions that are potentially overlapping.

An ultimate goal of physical mapping procedures is to obtain a complete contig for each type of chromosome within a full set. For example, in the case of humans, a complete physical map requires a contig for each of the 22 autosomes and for the X and Y chromosomes. Geneticists can also correlate cloned DNA fragments in a contig with locations along a chromosome

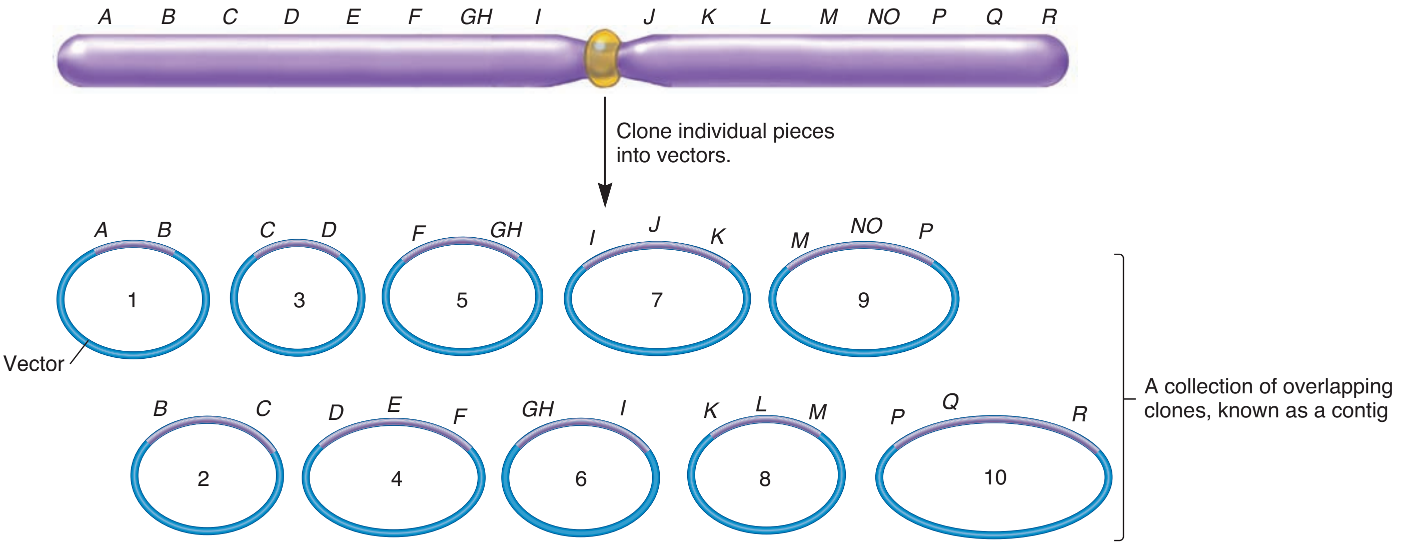

A B C D E F GH I J K L M NO P Q R
Clone individual pieces into vectors.
A B 1
C D 3
F GH 5
I J K 7
M NO P 9
Vector
B C 2
D E F 4
GH I 6
K L M 8
P Q R 10
A collection of overlapping clones, known as a contig

FIGURE 21.7 **The construction of a contig.** Large pieces of chromosomal DNA are cloned into vectors, and their order is determined by the identification of overlapping regions. For example, clone 3 ends with gene *D*, and clone 4 begins with gene *D*; clone 4 ends with gene *F*, and clone 5 begins with gene *F*.

Concept check: *What is a contig?*

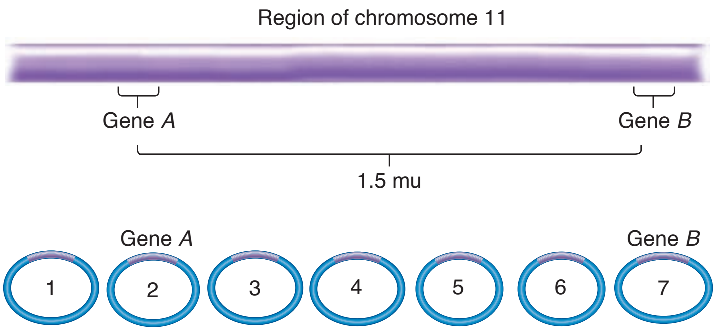


FIGURE 21.8 **The use of genetic markers to align a contig.** In this example, gene *A* and gene *B* had been mapped previously to specific regions of chromosome 11. Gene *A* was found within the insert of clone 2, gene *B* within the insert of clone 7. This made it possible to align the contig using gene *A* and gene *B* as genetic markers (i.e., reference points) along chromosome 11.

obtained from linkage or cytogenetic mapping. For example, a member of a contig may contain a gene, RFLP, or STS that was previously mapped by linkage analysis. Figure 21.8 considers a situation in which two members of a contig carry genes previously mapped by linkage analysis to be approximately 1.5 mu apart on chromosome 11. In this example, clone 2 has an insert that carries gene *A*, and clone 7 has an insert that carries gene *B*. Because a contig is composed of overlapping members, a researcher can align the contig along chromosome 11, starting with gene *A* and gene *B* as reference points. In this example, genes *A* and *B* serve as markers that identify the location of specific clones within the contig.

YAC, BAC, and PAC Vectors Are Used to Clone Large Segments of DNA

For large eukaryotic genomes, researchers often begin with vectors that can accept chromosomal DNA inserts of very large size. By having large insert sizes, a contig is more easily constructed and aligned because fewer recombinant vectors are needed. In general, most plasmid and viral vectors can accommodate inserts only a few thousand to perhaps tens of thousands of nucleotides in length. If a plasmid or viral vector has a DNA insert that is too large, it will have difficulty with DNA replication and is likely to suffer deletions in the insert.

By comparison, other cloning vectors, known as **artificial chromosomes,** can accommodate much larger sizes of DNA inserts. As their name suggests, they behave like chromosomes when inside of living cells. The first type to be made was the **yeast artificial chromosome (YAC),** which was developed by David Burke, Georges Carle, and Maynard Olson in 1987. An insert within a YAC can be several hundred thousand to about 2 million bp in length. For an average human chromosome, a few hundred YACs are sufficient to create a contig with fragments that span the entire length of the chromosome. By comparison, it would take thousands or even tens of thousands of recombinant plasmid vectors to create such a contig.

At the molecular level, YACs have structural similarities to normal eukaryotic chromosomes, yet have characteristics that make them suitable for cloning. The general structure of a YAC vector is shown at the top left of Figure 21.9. The YAC vector contains two telomeres (*TEL*), a centromere (*CEN*), a bacterial origin of replication (*ORI*), a yeast origin of replication (known as an *ARS,* for autonomous replication sequence), selectable markers, and unique cloning sites that are each recognized by a single restriction enzyme. Without an insert, the circular form of this vector can replicate in *E. coli*. After a large fragment has been inserted, the linear form of the vector can replicate in yeast.

In the experiment shown in Figure 21.9, chromosomal DNA is digested with the restriction enzyme *Eco*RI at a low concentration so only some of the restriction sites are cut. This partial digestion results in only occasional cleavage of the chromosomal DNA to yield very large DNA fragments.

The circular YAC vector is also digested with *Eco*RI and a second restriction enzyme, *Bam*HI, to yield two arms of the YAC. The YAC arms are then ligated to the large fragments of chromosomal DNA and transformed into yeast cells. When ligation occurs in the desired way, a large piece of chromosomal DNA becomes ligated to both arms of the YAC. Because each arm contains a different selectable marker, it is possible to select for the growth of yeast cells that carry a YAC construct having both arms.

Newer types of cloning vectors called **bacterial artificial chromosomes (BACs)** and **P1 artificial chromosomes (PACs)** have been constructed. BACs were developed from bacterial F factors, which are described in Chapter 9, and PACs were developed from P1 bacteriophage chromosomes. BACs and PACs typically can contain inserts up to 300,000 bp and sometimes larger. These vectors are somewhat easier to use than YACs because the DNA is inserted into a circular molecule and transformed into *E. coli*. BACs and PACs are more commonly used than YACs for the cloning of large DNA fragments.

Figure 21.10 shows a simplified drawing of a BAC cloning vector. The vector contains several genes that function in vector replication and segregation. The origin of replication is designated *oriS*, and the *repE* gene encodes a protein essential for replication at *oriS*. The *parA*, *parB*, and *parC* genes encode proteins required for the proper segregation of the vector into daughter cells. A chloramphenicol resistance gene, cm^R, provides a way to select for cells that have taken up the vector based on their ability to grow in the presence of the antibiotic chloramphenicol. The vector also contains unique restriction enzyme sites, such as *Hin*dIII, *Bam*HI, and *Sph*I, for the insertion of large fragments of DNA. These sites are located within the *lacZ* gene, which encodes the enzyme β–galactosidase. Vectors with DNA inserts can be determined by plating cells on a medium containing the compound X-Gal (as described for plasmid vectors in Chapter 19, Figure 19.2).

YAC, BAC, and PAC cloning vectors have been very useful in the construction of contigs that span long segments of chromosomes. They have been used as the first step in creating a rough physical map of a genome. Although this is an important step in physical mapping, the large insert sizes make them difficult to use in gene-cloning and sequencing experiments. Therefore, libraries containing recombinant vectors with smaller insert sizes are

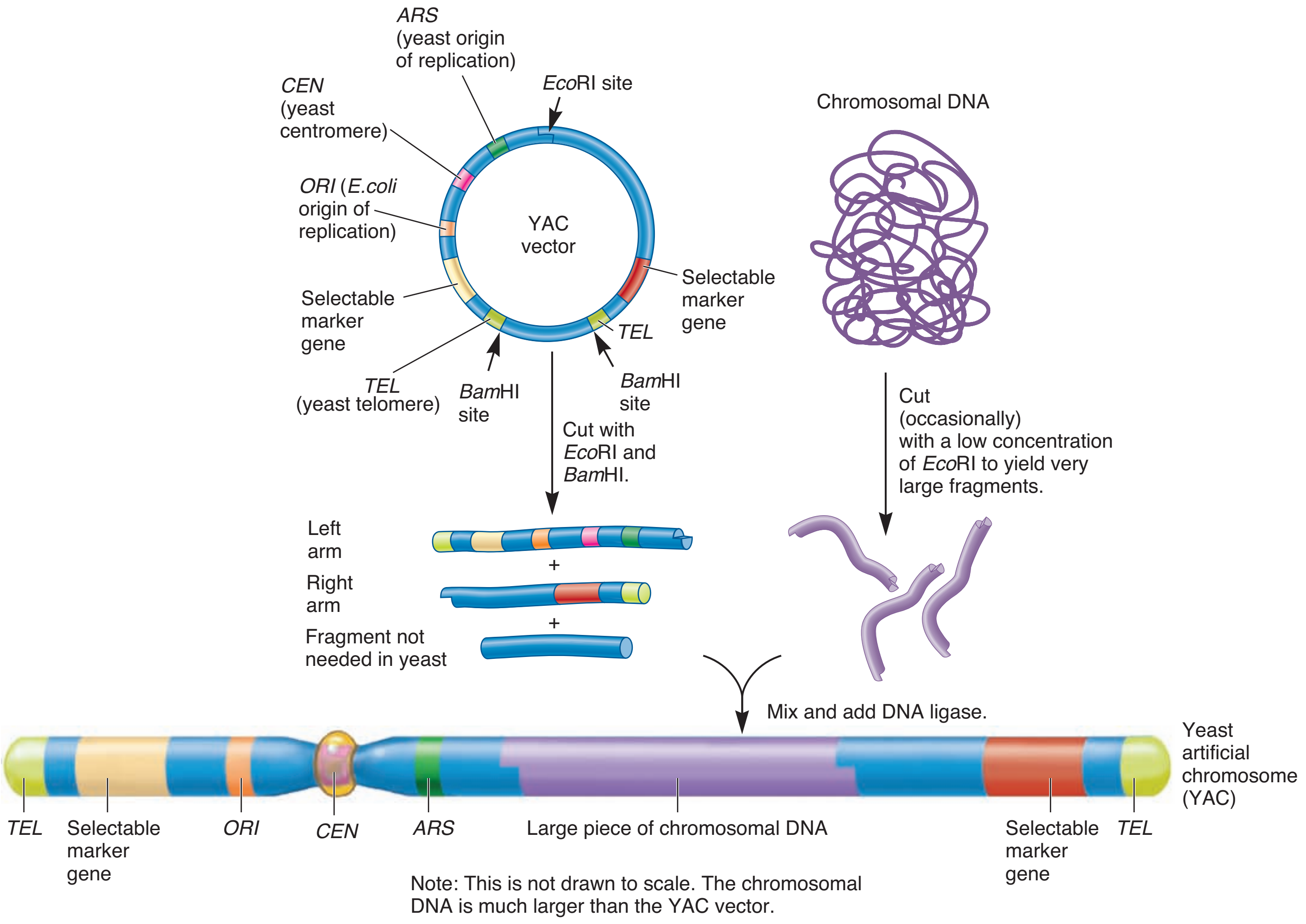

ARS
(yeast origin
of replication)
CEN
(yeast
centromere)
EcoRI site
ORI (E.coli
origin of
replication)
YAC
vector
Selectable
marker
gene
Selectable
marker
gene
TEL
TEL
(yeast telomere)
BamHI
site
BamHI
site
Cut with
EcoRI and
BamHI.
Chromosomal DNA
Cut
(occasionally)
with a low concentration
of EcoRI to yield very
large fragments.
Left
arm
+
Right
arm
+
Fragment not
needed in yeast
Mix and add DNA ligase.
Yeast
artificial
chromosome
(YAC)
TEL
Selectable
marker
gene
ORI
CEN
ARS
Large piece of chromosomal DNA
Selectable
marker
gene
TEL
Note: This is not drawn to scale. The chromosomal
DNA is much larger than the YAC vector.

FIGURE 21.9 **The use of a YAC vector in DNA cloning.** Note: *Eco*RI and *Bam*HI both cut DNA to yield sticky ends. In this diagram, only the EcoRI ends are shown with a notch to emphasize that these are the sticky ends that bind to the chromosomal DNA.

needed. In many cases, a type of cloning vector called a cosmid is used. A **cosmid** is a hybrid between a plasmid vector and phage λ; its DNA can replicate in a cell like a plasmid or be packaged into a protein coat like a phage. Cosmid vectors typically can accept DNA fragments that are tens of thousands of base pairs in length.

Researchers Can Make Genetic Maps of Large Eukaryotic Chromosomes

Figure 21.11 illustrates a comparison between cytogenetic, linkage, and physical maps of human chromosome 16. Actually, this is a very simplified map of chromosome 16. A much more detailed map is available, although it would take well over 10 pages of this textbook to print it! The top of this

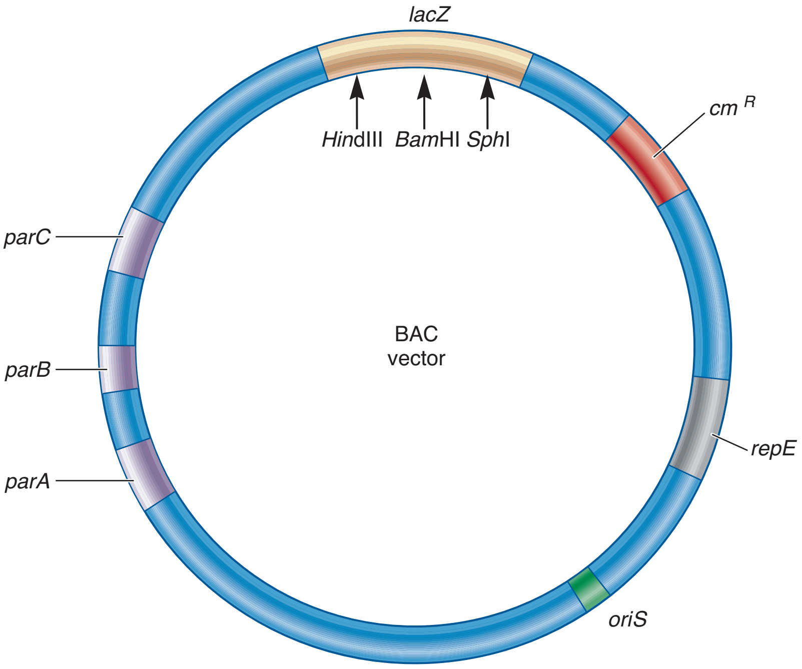

lacZ
HindIII
BamHI
SphI
cm R
parC
BAC
vector
parB
repE
parA
oriS

FIGURE 21.10 **A BAC vector.**

Concept check: *What is the main advantage of YACs, BACs, and PACs?*

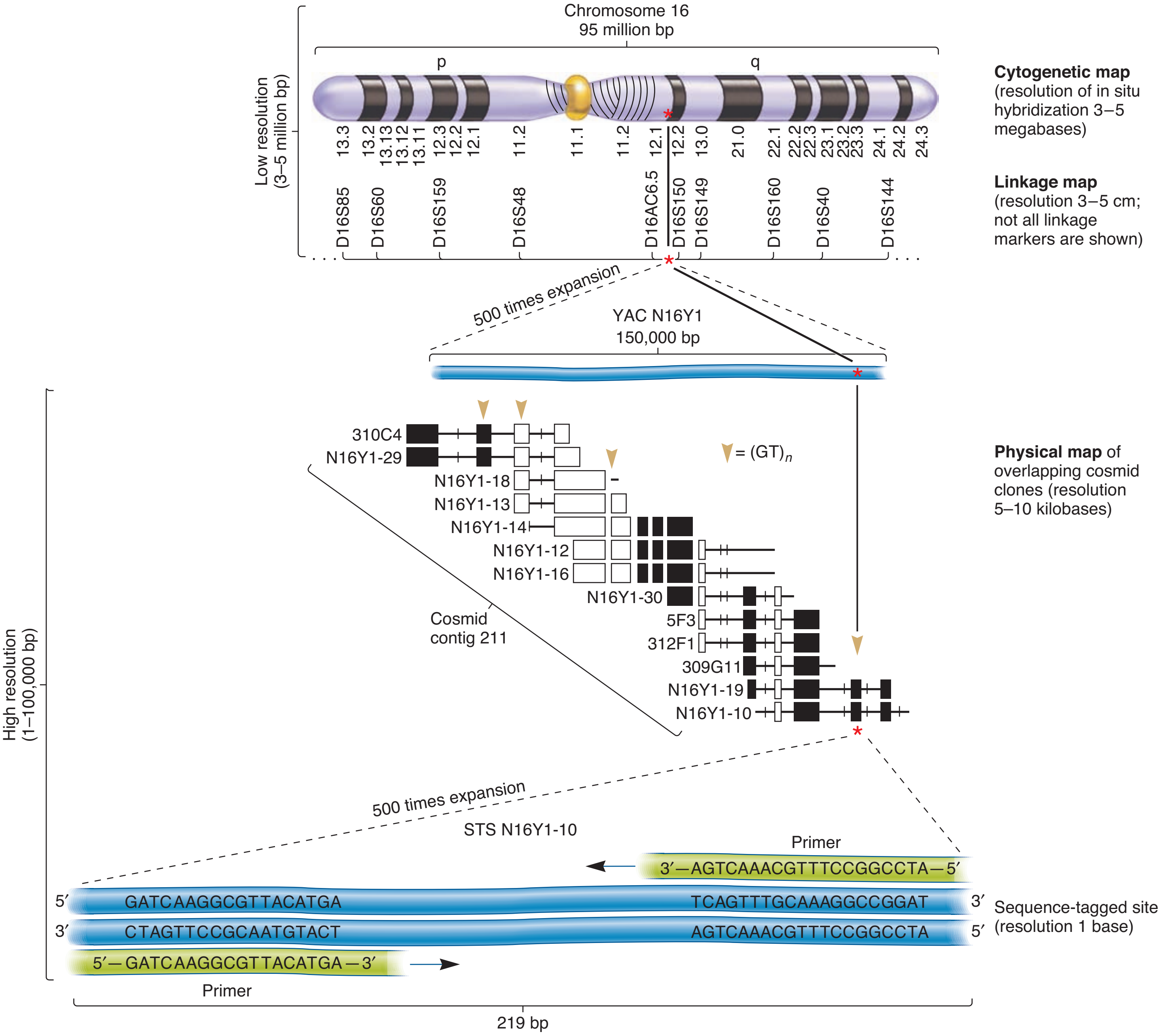

Chromosome 16
95 million bp
p
q
Low resolution (3–5 million bp)
13.3
13.2
13.13
13.12
13.11
12.3
12.2
12.1
11.2
11.1
11.2
12.1
12.2
13.0
21.0
22.1
22.2
22.3
23.1
23.2
23.3
24.1
24.2
24.3
D16S85
D16S60
D16S159
D16S48
D16AC6.5
D16S150
D16S149
D16S160
D16S40
D16S144
Cytogenetic map (resolution of in situ hybridization 3–5 megabases)
Linkage map (resolution 3–5 cm; not all linkage markers are shown)
500 times expansion
YAC N16Y1
150,000 bp
High resolution (1–100,000 bp)
310C4
N16Y1-29
N16Y1-18
N16Y1-13
N16Y1-14
N16Y1-12
N16Y1-16
N16Y1-30
5F3
312F1
309G11
N16Y1-19
N16Y1-10
= (GT)n
Cosmid contig 211
Physical map of overlapping cosmid clones (resolution 5–10 kilobases)
500 times expansion
STS N16Y1-10
Primer
3′—AGTCAAACGTTTCCGGCCTA—5′
5′
GATCAAGGCGTTACATGA
TCAGTTTGCAAAGGCCGGAT
3′
3′
CTAGTTCCGCAATGTACT
AGTCAAACGTTTCCGGCCTA
5′
5′—GATCAAGGCGTTACATGA—3′
Primer
219 bp
Sequence-tagged site (resolution 1 base)

FIGURE 21.11 **A correlation of cytogenetic, linkage, and physical maps of human chromosome 16.** The top part shows the cytogenetic map of human chromosome 16 according to its G banding pattern. A very simple linkage map of molecular markers (D16S85, D16S60, etc.) is aligned below the cytogenetic map. A correlation between the linkage map and a segment of the physical map is shown below the linkage map. A YAC clone designated YAC N16Y1 is located between markers D16AC6.5 and D16S150 on the linkage map. Pieces of DNA from this YAC were subcloned into cosmid vectors, and the cosmids (310C4, N16Y1-29, N16Y1-18, etc.) were aligned relative to each other. One of the cosmids (N16Y1-10) was sequenced, and this sequence was used to generate an STS shown at the bottom of the figure.

figure shows the banding pattern of this chromosome. Underneath the banded chromosome are molecular markers that have been mapped by linkage analysis. These same markers have been mapped cytogenetically. A complete contig of this chromosome has also been produced by generating a series of overlapping YACs. However, Figure 21.11 shows the location of only one YAC within the contig. Cosmids within this region are shown below the YAC. In addition, an STS is found within the region and provides a molecular marker for cosmids N16Y1-19 and N16Y1-10, as well as YAC N16Y1.

Positional Cloning Can Be Achieved Via Chromosome Walking

The creation of a contig bears many similarities to a gene-cloning strategy known as **positional cloning,** in which a gene is cloned based on its mapped position along a chromosome. This approach has been successful in the cloning of many human genes, particularly those that cause genetic diseases when mutated. These include genes involved in cystic fibrosis, Huntington disease, and Duchenne muscular dystrophy.

One method of positional cloning is known as **chromosome walking.** To initiate this type of experiment, a gene's position relative to a marker must be known from mapping studies. For example, a gene may be known to be fairly close to a previously mapped gene or RFLP marker. This provides a starting point to molecularly "walk" toward the gene of interest.

Figure 21.12 considers a chromosome walk in which the goal is to locate a gene that we will call gene *A*. In this example, linkage mapping studies have revealed that gene *A* is relatively close to another gene, called gene *B*, that was previously cloned. Gene *A* and gene *B* have been deduced from genetic crosses to be approximately 1 mu apart. To begin this chromosome walk, a cloned DNA fragment that contains gene *B* and flanking sequences can be used as a starting point.

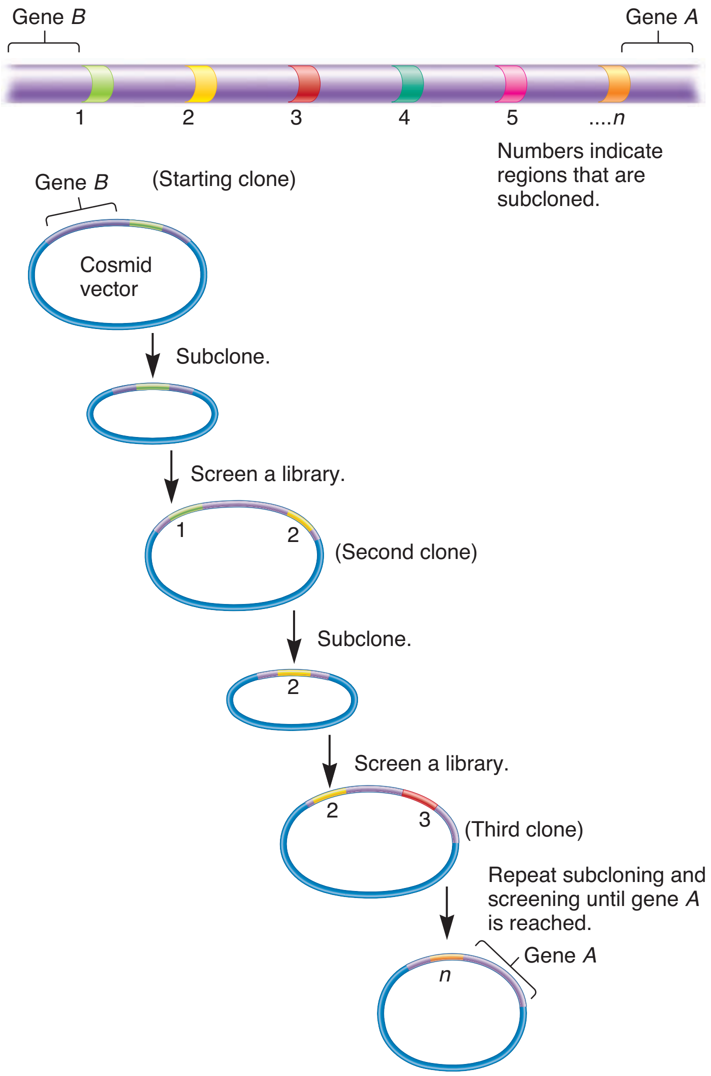


FIGURE 21.12 **The technique of chromosome walking.**

Concept check: Why is this technique used?

To walk from gene *B* to gene *A*, a series of library screening methods are followed. In this example, the starting materials are a cosmid library and a clone containing gene *B*. A small piece of DNA from the first cosmid vector containing gene *B* is inserted into another vector. This is called **subcloning.** The subcloned DNA is labeled and used as a probe to screen a cosmid library. This enables the researchers to identify a second clone that extends into the region that is closer to gene *A*. A subclone from this second clone is then used to screen the library a second time. This allows the researchers to identify a clone that is even closer to gene *A*. This repeated pattern of subcloning and library screening is used to reach gene *A*. The term *chromosome walking* is an appropriate description of this technique, because each clone takes you a step closer to the gene of interest. When starting at gene *B* in Figure 21.12, researchers also want to have markers to the left of gene *B* to ensure they were not walking in the wrong direction.

The number of steps required to reach the gene of interest depends on the distance between the starting and ending points and on the sizes of the DNA inserts in the library. If the two points are 1 mu apart, they are expected to be approximately 1 million bp apart, although the correlation between map units and physical distances can vary greatly. In a typical walking experiment, each clone might have an average insert size of 50,000 bp. Therefore, it takes about 20 walking steps to reach the gene of interest. Researchers want to locate starting points in a chromosome walking experiment that are as close as possible to the gene they wish to identify.

How do researchers know when they have reached a gene of interest? In the case of a gene that causes a disease when mutant, researchers conduct their walking steps on DNA from both an unaffected and an affected individual. Each set of clones is subjected to DNA sequencing, and those DNA sequences are compared with each other. When the researchers reach a spot where the DNA sequences differ between the unaffected and the affected individual, such a site may be within the gene of interest. However, this has to be confirmed by sequencing the region from several unaffected and affected individuals to be certain the change in DNA sequence is correlated with the disease.

21.4 REVIEWING THE KEY CONCEPTS

- A physical map of a chromosome is made by producing a contiguous collection of clones called a contig, which spans an entire chromosome (see Figures 21.7, 21.8).
- YACs, BACs, PACs, and cosmids are vectors that can carry a long segment of DNA and are used to make contigs (see Figures 21.9–21.11).
- Chromosome walking is a technique for cloning a gene that is close to a known marker (see Figure 21.12).

21.4 COMPREHENSION QUESTIONS

1. What is a contig?
   a. A fragment of DNA that has been inserted into a vector
   b. A series of vectors that contain inserts that have overlapping regions of chromosomal DNA
   c. A method of identifying a disease-causing allele
   d. A method of sequencing DNA

2. A vector that can carry a large fragment of chromosomal DNA is a
   a. YAC.
   b. BAC.
   c. PAC.
   d. All of the above
3. Chromosomal walking is a type of ______________ in which a researcher begins at a specific site on a chromosome and analyzes __________________ until the gene of interest is reached.
   a. DNA sequencing, a series of subclones
   b. positional cloning, a series of subclones
   c. DNA sequencing, bands on a gel
   d. positional cloning, bands on a gel

21.5 GENOME-SEQUENCING PROJECTS

Genome-sequencing projects are research endeavors that have the ultimate goal of determining the sequence of DNA bases of the entire genome of a given species. Such projects involve many participants, including scientists who isolate DNA and perform DNA-sequencing reactions, as well as theoreticians who gather the DNA sequence information and assemble it into a long DNA sequence for each chromosome. For bacteria and archaea, which usually have just one chromosome, the genome sequence is that of a single chromosome. For eukaryotes, each chromosome must be sequenced. For humans, the genome sequence consists of sequences of 22 autosomes, 2 sex chromosomes, and the mitochondrial genome.

In just a couple of decades, our ability to map and sequence genomes has improved dramatically. In 1995, the first genome to be entirely sequenced was that of the bacterium *Haemophilus influenzae*. This bacterium causes a variety of diseases in humans, including respiratory illnesses and bacterial meningitis. *H. influenzae* has a relatively small genome consisting of approximately 1.8 Mb of DNA in a single circular chromosome. As of 2011, the complete genome sequences have been obtained from hundreds of different species, including over 300 prokaryotes and 100 eukaryotes. Considering that the first genome sequence was generated in 1995, the progress of genome sequencing projects since then has been truly remarkable! In this section, we examine the approaches that researchers follow when tackling such a large project. We will also survey some of the general goals of the Human Genome Project, the largest project of its kind, and compare the results from the genome sequencing of various species.

Shotgun Sequencing Has Followed Different Strategies

When sequencing an entire genome, researchers must consider factors such as genome size, the efficiency of the methods used to sequence DNA, and the costs of the project. Since genome-sequencing projects began in the 1990s, researchers have learned that the most efficient and inexpensive way to sequence genomes is via an approach called **shotgun sequencing,** in which DNA fragments to be sequenced are randomly generated from larger DNA fragments. In this method, genomic DNA is isolated and broken into smaller fragments, typically 1500 bp or longer. Until recently, the researchers then used the technique of dideoxy sequencing, described in Chapter 19, to randomly sequence fragments from the genome. (Newer methods of DNA sequencing are described later in this chapter.) As a matter of chance, some of the fragments overlap, as shown schematically in **Figure 21.13.** The DNA sequence in two different fragments is identical in the overlapping region. This allows researchers to order them as they are found in the intact chromosome. An advantage of shotgun DNA sequencing is that it does not require extensive mapping, which can be very time-consuming and expensive. A disadvantage is that researchers waste some time sequencing the same region of DNA more times than needed.

For many genome-sequencing projects in the past, researchers typically followed two types of shotgun strategies: hierarchical shotgun sequencing or whole-genome shotgun sequencing. In the late 1990s, when researchers first began to sequence the larger genomes from eukaryotic species, they first took the approach of **hierarchical shotgun sequencing** (**Figure 21.14a**). To begin this process, researchers cloned large (e.g., 150,000 bp long) DNA fragments into BACs (or YACs) to generate a contig for each chromosome. For DNA sequencing, researchers chose a set of clones from each chromosome contig that had relatively short overlapping regions. Each BAC clone in a chosen set was then subjected to shotgun sequencing. A clone was broken into smaller pieces (e.g., 2000 bp), and those small pieces were cloned into vectors. The recombinant vectors were then randomly sequenced using a primer that annealed at a site next to the inserted DNA (see Figure 19.17). The method of dideoxy sequencing yielded a DNA sequence, usually 500 to 1000 bp in length, at one end of each insert. The resulting DNA sequences were analyzed by a computer program that identified overlapping sequences between two DNA fragments and assembled the DNA sequence into one long sequence along a single chromosome.

In the hierarchical method, the nucleotide sequence was determined clone by clone. For this reason, the method is also called the clone-by-clone, or BAC-to-BAC, approach. This strategy had the advantage of making it easier to align the overlapping sequences because the DNA sequences were grouped according to the BAC clone from which the sequence was obtained. Another issue is that this method minimized the amount of DNA sequencing by reducing the amount of sequencing in overlapping

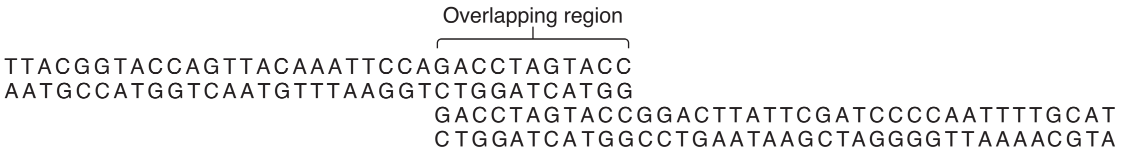
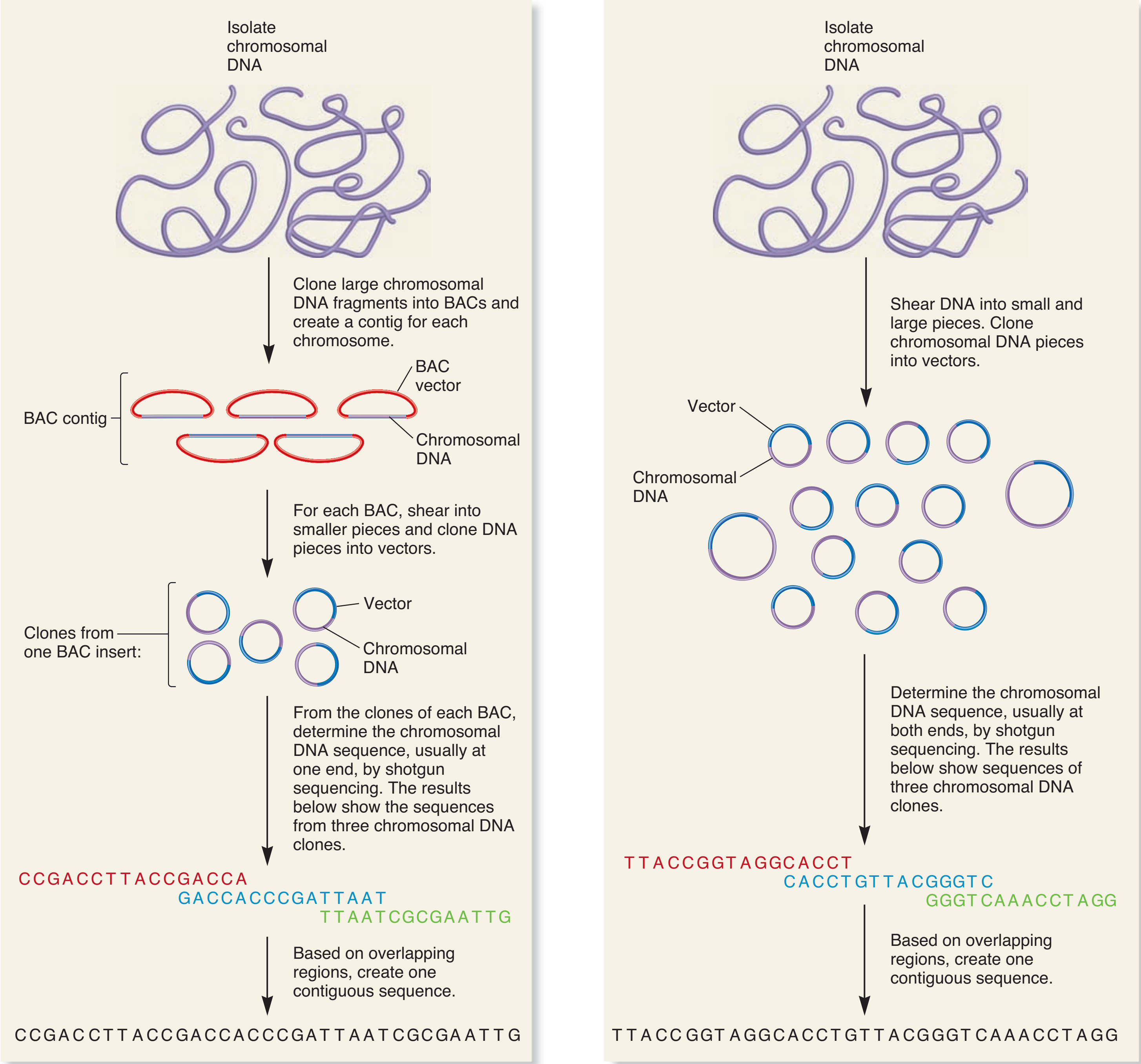

Overlapping region
TTACGGTACCAGTTACAAATTCCAGACCTAGTACC
AATGCCATGGTCAATGTTTAAGGTCTGGATCATGG
GACCTAGTACCGGACTTATTCGATCCCCAATTTTGCAT
CTGGATCATGGCCTGAATAAGCTAGGGGTTAAAACGTA

FIGURE 21.13 **A comparison of two DNA fragments that contain an overlapping region.**

regions. This was an advantage in the 1990s when DNA sequencing was an expensive and labor-intensive procedure. However, a disadvantage of this approach is that BAC mapping can be time-consuming and expensive.

In more recent years, advances in DNA-sequencing technology greatly diminished the time and expense of dideoxy sequencing. Genome-sequencing projects in the 1990s through first decade of the 21st century more commonly followed a strategy called **whole-genome shotgun sequencing.** In this method, the physical mapping step used in the hierarchical approach was bypassed. The DNA from the entire genome was isolated and sheared into small (e.g., 2000- or 10,000-bp) and large (50,000- or 150,000-bp) fragments and cloned into vectors (**Figure 21.14b**). Each insert was usually sequenced at both ends—an approach called **double-barrel shotgun sequencing.** As in the hierarchical approach, the DNA sequences, which


Isolate chromosomal DNA
Clone large chromosomal DNA fragments into BACs and create a contig for each chromosome.
BAC vector
BAC contig
Chromosomal DNA
For each BAC, shear into smaller pieces and clone DNA pieces into vectors.
Vector
Clones from one BAC insert:
Chromosomal DNA
From the clones of each BAC, determine the chromosomal DNA sequence, usually at one end, by shotgun sequencing. The results below show the sequences from three chromosomal DNA clones.
CCGACCTTACCGACCA
GACCACCCGATTAAT
TTAATCGCGAATTG
Based on overlapping regions, create one contiguous sequence.
CCGACCTTACCGACCACCCGATTAATCGCGAATTG
Isolate chromosomal DNA
Shear DNA into small and large pieces. Clone chromosomal DNA pieces into vectors.
Vector
Chromosomal DNA
Determine the chromosomal DNA sequence, usually at both ends, by shotgun sequencing. The results below show sequences of three chromosomal DNA clones.
TTACCGGTAGGCACCT
CACCTGTTACGGGTC
GGGTCAAACCTAGG
Based on overlapping regions, create one contiguous sequence.
TTACCGGTAGGCACCTGTTACGGGTCAAACCTAGG

(a) Hierarchical genome shotgun sequencing

(b) Whole-genome shotgun sequencing

FIGURE 21.14 **Two approaches of shotgun sequencing of genomes.** Note: The BAC vector and chromosomal DNA are not drawn to scale. The chromosomal DNA would be much longer than the BAC vector DNA.

Concept check: *Which of these two procedures is now more commonly done?*

were 500 bp to 1000 bp in length, were then analyzed by a computer program that identified overlapping regions between two DNA fragments and assembled the data into one long sequence. The double-barrel strategy improved the ability of researchers to align their DNA sequences. Most of the DNA sequencing in a whole-genome shotgun approach was carried out on small DNA fragments. However, due to the prevalence of repetitive sequences, particularly within the genomes of eukaryotic species, the alignment of small pieces sometimes was difficult. The sequencing of both ends of longer fragments made alignment easier, because the relative spacing of particular DNA sequences was known.

The Human Genome Project Was the Largest Genome-Sequencing Project in History

Due to its large size, the sequencing of the human genome was an enormous undertaking. Scientists had been discussing how to undertake this project since the mid-1980s. In 1988, the National Institutes of Health (NIH) established an Office of Human Genome Research, with James Watson as its first director. The **Human Genome Project,** which officially began on October 1, 1990, was a 13-year effort coordinated by the U.S. Department of Energy and the NIH. From its outset, the Human Genome Project had the following goals:

1. *To obtain a linkage map of the human genome.* This has involved the identification of thousands of genetic markers and their localization along the autosomes and sex chromosomes.
2. *To obtain a physical map of the human genome.* This has required the cloning of many segments of chromosomal DNA into BACs, YACs, and cosmids.
3. *To obtain the DNA sequence of the entire human genome.* The first (nearly complete) sequence was published in February 2001. This was considered a first draft. A second draft was published in 2003, and the completed maps and sequences for all of the human chromosomes were published by 2006. The entire genome is approximately 3 billion bp in length. If the entire human genome were typed in a textbook like this, it would be nearly 1 million pages long!
4. *To develop technology for the management of human genome information.* The amount of information obtained from this project is staggering, to say the least. The Human Genome Project developed user-friendly tools to provide scientists easy access to up-to-date information obtained from the project. The Human Genome Project also developed analytical tools for interpreting genome information.
5. *To analyze the genomes of other model organisms.* These include bacterial species (e.g., *Escherichia coli* and *Bacillus subtilis*), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (a nematode), *Arabidopsis thaliana* (a flowering plant), and *Mus musculus* (mouse).
6. *To develop programs focused on understanding and addressing the ethical, legal, and social implications of the results obtained from the Human Genome Project.* The Human Genome Project has sought to identify the major genetic issues that will affect members of our society and to develop policies to address these issues. For example, what is an individual's right to privacy regarding genetic information? Some people are worried that their medical insurance company may discriminate against them if it is found that they carry a disease-causing or otherwise deleterious gene.
7. *To develop technological advances in genetic methodologies.* Some of the efforts of the Human Genome Project have involved improvements in molecular genetic technology such as gene cloning, contig construction, and DNA sequencing. The project has also been aimed at developing computer technology for data processing, storage, and the analysis of sequence information.

A great benefit expected from the characterization of the human genome is the ability to identify and study the sequences of our genes. Mutations in many different genes are known to be correlated with human diseases, which include cancer, heart disease, and many other abnormalities. The identification of mutant genes that cause inherited diseases was a strong motivation for the Human Genome Project. A detailed genetic and physical map has made it profoundly easier for researchers to locate such genes. Furthermore, a complete DNA sequence of the human genome provides researchers with insight into the types of proteins encoded by these genes. The cloning and sequencing of disease-causing alleles is expected to play an increasingly important role in the diagnosis and treatment of disease.

Innovations in DNA Sequencing Have Made It Faster and Less Expensive

Since it was invented in the 1980s, technological advances in DNA sequencing have been aimed at making it faster and less expensive. The Human Genome Project, which began in the early 1990s, was originally estimated to cost about $3 billion to sequence a single genome. However, cost reductions due to innovations in DNA-sequencing technology drove the actual cost down to about $300 million. By the end of the project, researchers estimated that if they were starting again, they could have sequenced the genome for less than $50 million. The project took about 13 years to complete. In 2007, researchers undertook the sequencing of James Watson's genome, which cost less than $1 million. By 2010, the cost of sequencing of a human genome had been reduced to approximately $10,000 to $100,000. By 2015, some scientists predict that the sequencing of a single human genome will cost less than $1000 and take only a few days to complete. Such innovation will make it feasible to sequence an individual's genome as a routine diagnostic procedure.

The ability to rapidly sequence large amounts of DNA is often referred to as **high-throughput sequencing.** Different types of technological advances have made this possible. First, different aspects of DNA sequencing have become automated so that samples can be processed rapidly in a machine. For example, in Chapter 19, we considered how fluorescently labeled nucleotides can automate the ability to read a DNA sequence by using a fluorescence detector. A second advance from high-throughput

TABLE **21.2**

Examples of Next-Generation DNA-Sequencing Technologies

Technology*	DNA Preparation	Enzyme(s) Used	Detection
Second-Generation DNA Sequencers			
Roche 454/FLX Pyrosequencer	DNA fragments are bound to small beads, which are dropped into tiny wells in a fiber-optic chip.	DNA polymerase, ATP sulfurylase, luciferase, apyrase	Pyrophosphate release activates luciferase, which gives off light.
Illumina/Solexa Genome Analyzer	DNA fragments are bound to a flow cell surface.	DNA polymerase	Four different fluorescently labeled nucleotides are detected.
Third-Generation DNA Sequencers			
Pacific Biosciences SMRT	DNA fragments and DNA polymerase are trapped within tiny holes on a thin metal film.	DNA polymerase	The growth of individual DNA molecules is monitored by fluorescence imaging.
Helicos Biosciences tSMS	DNA fragments are bound to a flow cell surface.	DNA polymerase	The growth of individual DNA molecules is monitored by fluorescence imaging.
ZS Genetics TEM	DNA is labeled with heavier elements, such as iodine or bromine.	None	The DNA sequence of single molecules is read via transmission electron microscopy (TEM).

*Also includes the company name associated with the technology.

sequencing technologies is parallel sequencing. This means that multiple samples can be processed at once. The first sequencing machines, also called platforms, could simultaneously perform many sequencing runs via multiple capillary gels. For example, DNA-sequencing machines produced by Applied Biosystems, which rely on dideoxy sequencing, run 96 capillaries in parallel. Each capillary gel is capable of producing between 500 and 1000 bases of DNA sequence.

Although the dideoxy sequencing method is still in use, newer high-throughput platforms are based on different methods of sequencing DNA. Table 21.2 describes a few of these methods, which are often referred to as **next-generation sequencing technologies** because they have superseded the dideoxy method for large sequencing projects. What sets next-generation sequencers apart from conventional capillary-based sequencing? One key technological advance is the ability to process thousands or even millions of sequence reads in parallel rather than only 96 at a time. This massively parallel throughput may require only one or two instrument runs to complete the sequencing of an entire prokaryotic genome. Also, next-generation sequencers are able to use samples that contain mixtures of DNA fragments that have not been subjected to the conventional vector-based cloning. By comparison, shotgun methods of DNA sequencing using capillary sequencing involve DNA-cloning steps (see Figure 21.14). The elimination of such cloning steps saves a great deal of time and money.

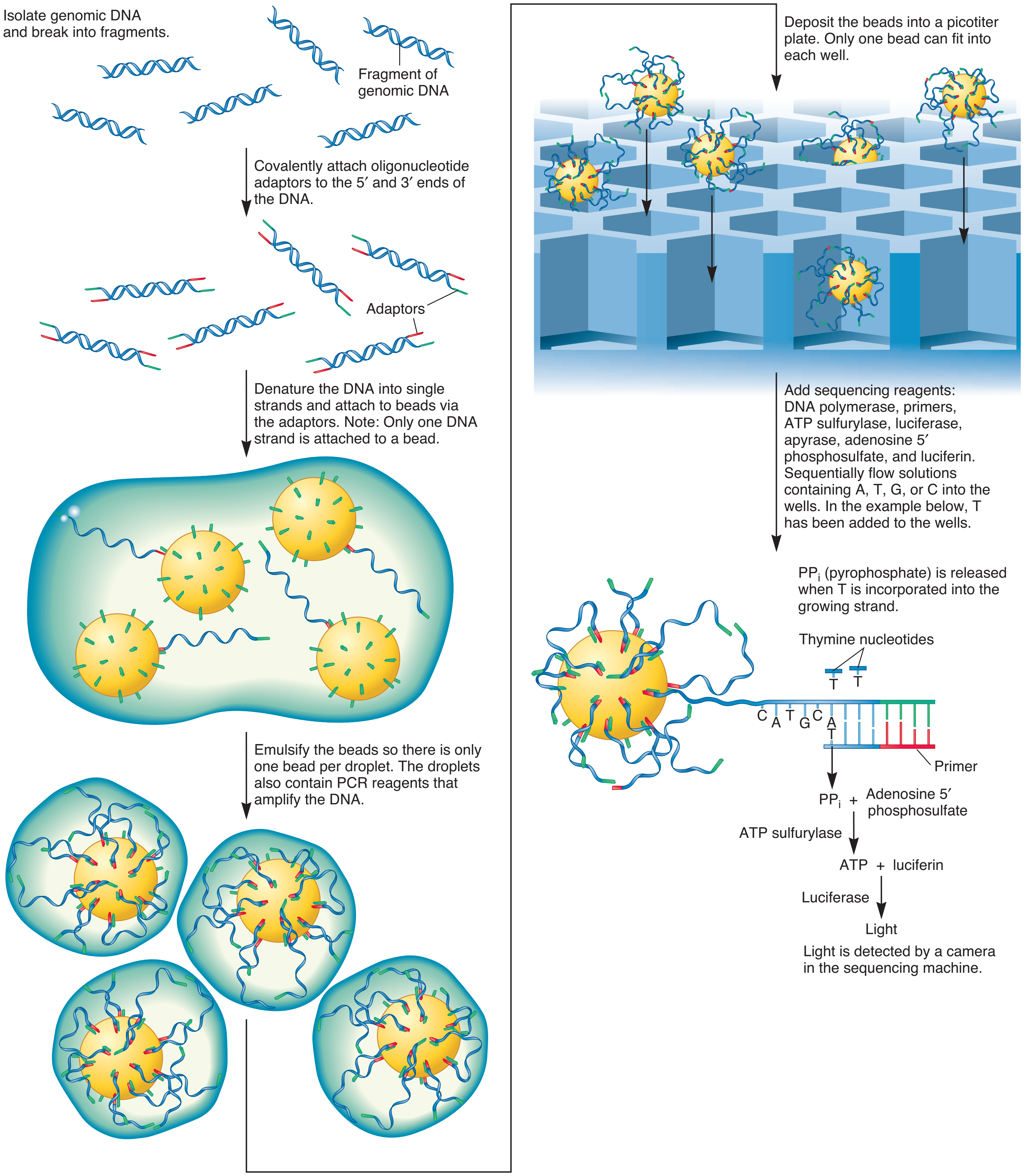
The newer sequencing platforms employ a complex interplay of enzymology, chemistry, high-resolution optics, and new approaches to processing the data. These instruments allow for easy sample preparation steps prior to DNA sequencing. Most of them involve strategies in which fragmented DNA is immobilized in a fixed position and repeatedly exposed to reagents. The newer sequencing methods, called second-generation sequencing platforms, use PCR to amplify the DNA, whereas third-generation sequencers actually read single DNA molecules.

As an example of next-generation sequencing, Figure 21.15 considers the technology called **pyrosequencing,** which relies on the release of pyrophosphate. This method was developed by Pal Nyren and Mostafa Ronaghi in 1996 and is the basis for the Roche 454/FLX Pyrosequencer. Samples, such as genomic DNA, are broken into small 300- to 800-bp fragments. Short oligonucleotides called adaptors are then linked to the 5′ and 3′ ends of the DNA fragments. The adaptors are used for the purification, PCR amplification, and DNA-sequencing steps. The DNA is then denatured into single strands. The mixture of single-stranded DNA fragments with adaptors is called the sample library.

The single-stranded DNA fragments are then attached to beads via the adaptors. Initially, just one DNA strand is attached per bead. The beads are emulsified in an oil-water mixture such that each bead becomes localized into a single droplet. The mixture also contains PCR reagents, including primers that are complementary to the adaptors. During this step, the single-stranded DNA on the bead becomes amplified into many identical products, which also become attached to the beads. At this stage, each bead contains about one million copies of a particular DNA segment.

Next, each bead is placed into a well of a picotiter plate; the diameter of each well is large enough to accommodate only one bead. Sequencing reagents, which include primers that are complementary to the adaptors, are added to the wells. The picotiter plate acts like a flow cell in which pure solutions of nucleotides are added in a stepwise fashion and allowed to flow over the DNA. The synthesis of DNA is monitored in real time. Therefore, this type of method is referred to as **sequencing by synthesis (SBS)** because it involves the identification of each nucleotide immediately after its incorporation into a DNA strand by DNA polymerase.

When a solution has been added that contains one particular nucleotide, how does the machine determine if that nucleotide has been added to a growing DNA strand? The pyrosequencing method relies on the functions of three other



Isolate genomic DNA and break into fragments.

Covalently attach oligonucleotide adaptors to the 5′ and 3′ ends of the DNA.

Denature the DNA into single strands and attach to beads via the adaptors. Note: Only one DNA strand is attached to a bead.

Emulsify the beads so there is only one bead per droplet. The droplets also contain PCR reagents that amplify the DNA.

Deposit the beads into a picotiter plate. Only one bead can fit into each well.

Add sequencing reagents: DNA polymerase, primers, ATP sulfurylase, luciferase, apyrase, adenosine 5′ phosphosulfate, and luciferin. Sequentially flow solutions containing A, T, G, or C into the wells. In the example below, T has been added to the wells.

PP_i (pyrophosphate) is released when T is incorporated into the growing strand.

PP_i + Adenosine 5′ phosphosulfate

ATP sulfurylase

ATP + luciferin

Luciferase

Light

Light is detected by a camera in the sequencing machine.

FIGURE 21.15 **Pyrosequencing, an example of a newer high-throughput sequencing technology.**

Concept check: *Is this a "sequencing by synthesis" method? Explain.*

enzymes, ATP sulfurylase, luciferase, and apyrase, along with the molecules adenosine 5′-phosphosulfate and luciferin, which are also included in the sequencing reaction. If a nucleotide has been incorporated into a DNA strand, pyrophosphate (PP_i) is released. ATP sulfurylase uses that PP_i along with adenosine 5′-phosphosulfate to make ATP. The ATP is then used by luciferase to break down luciferin. This reaction gives off light, which is detected by a camera in the sequencing machine. Therefore, light is given off only when a nucleotide with the correct base is incorporated into a DNA strand. Apyrase degrades unincorporated nucleotides and ATP, and the reaction can start again by flowing a new solution containing a particular nucleotide over the picotiter plate. By sequentially adding solutions with only one of the four possible nucleotides (A, T, G, or C), the sequence of the DNA strand can monitored in real time by determining when light is given off.

Many Genome Sequences Have Been Determined

The amazing advances in DNA-sequencing technology have enabled researchers to determine the complete genome sequences of hundreds of species. By 2011, over 300 prokaryotic and 100 eukaryotic genomes had been sequenced, including more than two dozen mammalian genomes.

How do researchers decide which genomes to sequence? Motivation behind genome-sequencing projects comes from a variety of sources. For example, basic research scientists can greatly benefit from a genome sequence. It allows them to know which genes a given species has, and it aids in the cloning and characterization of such genes. This has been the impetus for genome projects involving model organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, and *Mus musculus*. A second impetus for genome sequencing has involved human disease. As noted previously, researchers expect that the sequencing of the human genome will aid in the identification of genes that, when mutant, play a role in disease. Likewise, the decision to sequence many bacterial, fungal, and protist genomes has been related to the role of these species in infectious diseases. Hundreds of microbial genomes have been sequenced. Many of them are from species that are pathogenic in humans. The sequencing of such genomes may help to elucidate which genes play a role in the infection process.

In addition, the genomes of agriculturally important species have been the subject of genome sequencing. An understanding of a species' genome may aid in the development of new strains of livestock and plant species that have improved traits from an agricultural perspective. Finally, genome-sequencing projects help us to better understand the evolutionary relationships among living species. This approach, called **comparative genomics,** uses information from genome projects to understand the genetic variation among different populations. We will explore this topic in Chapter 27.

Table 21.3 describes the results of several genome-sequencing projects that have been completed. Newly completed genome sequences are emerging rapidly, particularly those of microbial species. As we obtain more genome sequences, it becomes progressively more interesting to compare them to each other as a way to understand the process of evolution. As we will discover in Chapter 22, the field of functional genomics enables researchers to probe the roles of many genes as they interact to generate the phenotypic traits of the species that contain them.

21.5 REVIEWING THE KEY CONCEPTS

- Shotgun DNA sequencing has been commonly used to sequence the genomes of many species. Overlapping regions allow researchers to determine the contiguous DNA sequence. Shotgun sequencing can follow a hierarchical or whole-genome shotgun sequencing approach (see Figures 21.13, 21.14).
- The Human Genome Project resulted in the successful sequencing of the entire human genome.
- Newer methods of DNA sequencing can process many samples of DNA simultaneously and are superseding the dideoxy method. An example is pyrosequencing (see Table 21.2, Figure 21.15).
- Since 1995, the genomes of hundreds of species have been sequenced (see Table 21.3).

21.5 COMPREHENSION QUESTIONS

1. Shotgun sequencing is a method of DNA sequencing in which
   a. the DNA fragments to be sequenced are randomly generated from larger DNA fragments.
   b. the sequencing reactions are carried out in rapid succession.
   c. the samples to be sequenced are rapidly generated by PCR.
   d. all of the above occur.
2. Which of the following was *not* a goal of the Human Genome Project?
   a. To obtain the DNA sequence of the entire human genome
   b. To successfully clone a mammal
   c. To develop technology for the management of human genome information
   d. To analyze the genomes of model organisms
3. A prokaryotic genome is about 4 million bp in length. About how many genes would you expect it to contain?
   a. 400
   b. 4000
   c. 40,000
   d. 400,000

21.6 TRANSPOSITION

Thus far, we have considered how genomes are analyzed and subjected to DNA sequencing. Genome sequencing projects have revealed that a sizable portion of many species' genomes are composed of repetitive sequences. In many cases, the repetitive sequences are due to **transposition,** which involves the integration of small segments of DNA into a chromosome. Transposition can occur at many different locations within the genome. The DNA segments that transpose themselves are known as **transposable elements (TEs).** TEs have sometimes been referred to as "jumping genes" because they are inherently mobile.

Transposable elements were first identified by Barbara McClintock in the early 1950s from her classic studies with corn

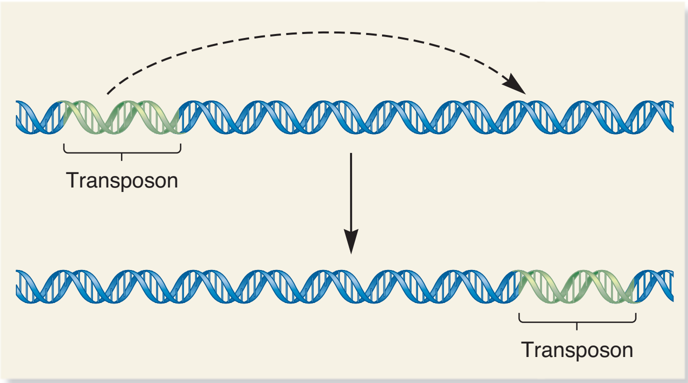
TABLE 21.3

Examples of Genomes That Have Been Sequenced

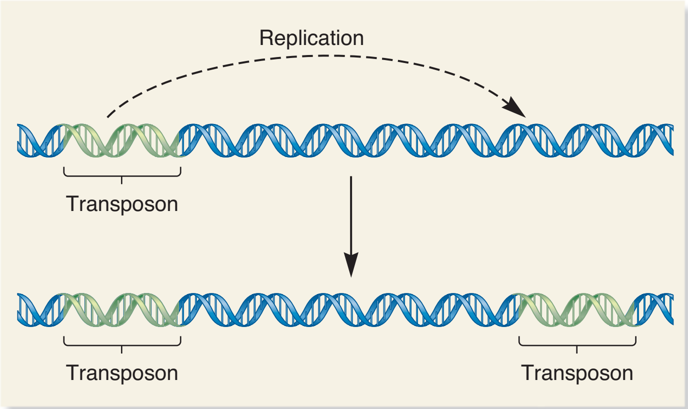
Species	Genome Size (bp)*	Number of Genes†	Description
Prokaryotic Genomes			
Bacteria			
Mycoplasma genitalium	580,000	521	Inhabitant of the human genital tract with a very small genome
Helicobacter pylori	1,668,000	1590	Common inhabitant of the stomach that may cause gastritis, peptic ulcer, and gastric cancer
Mycobacterium tuberculosis	4,412,000	4294	Bacterial species that causes tuberculosis
Escherichia coli	4,639,000	4289	Widespread bacterial inhabitant of the gut of animals; also a model research organism
Archaea			
Thermoplasma volcanium	1,580,000	1494	Archaea with an optimal growth temperature of 60°C and pH optimum of <2.0
Pyrococcus abyssi	1,760,000	1765	Originally isolated from samples taken close to a hot spring situated 3500 m deep in the southeast Pacific. Optimal growth conditions are 103°C and 200 atm
Sulfolobus solfataricus	2,990,000	2977	Found in terrestrial volcanic hot springs with optimum growth occurring at a temperature of 75–80°C and pH 2–3
Eukaryotic Genomes			
Protists			
Plasmodium falciparum	22,900,000	5268	Parasitic protist that causes malaria in humans
Entamoeba histolytica	23,800,000	9938	Amoeba that causes dysentery in humans
Fungi			
Saccharomyces cerevisiae (baker's yeast)	12,100,000	6294	Structurally simple eukaryotic species that has been extensively studied by researchers to understand eukaryotic genetics and cell biology
Neurospora crassa	40,000,000	10,082	Common bread mold and also a structurally simple eukaryotic species that has been extensively studied by researchers
Plants			
Arabidopsis thaliana	142,000,000	26,000	Model organism studied by plant biologists
Oryza sativa (rice)	440,000,000	40,000	Cereal grain with a relatively small genome. It is very important worldwide as a food crop.
Populus trichocarpa (Balsam poplar)	550,000,000	45,555	Tree with a relatively small genome
Animals			
Caenorhabditis elegans	97,000,000	19,000	Nematode worm that has been used as a model organism to study animal development
Drosophila melanogaster (fruit fly)	175,000,000	14,000	Used as a model organism to study many genetic phenomena, including development
Anopheles gambiae	278,000,000	13,683	Mosquito that carries the malaria parasite, *Plasmodium falciparum*
Canis familiaris (dog)	2,400,000,000	19,300	Common house pet
Mus musculus (mouse)	2,500,000,000	24,174	Rodent and a model organism studied by researchers
Pan troglodytes (chimpanzee)	3,100,000,000	25,000	Primate and the closest living relative to humans
Homo sapiens (human)	3,200,000,000	25,000	Primate

*In some cases, the values indicate the estimated amount of DNA. For eukaryotic genomes, DNA sequencing is considered completed in the euchromatic regions. The DNA in certain heterochromatic regions cannot be sequenced, and the total amount is difficult to estimate.

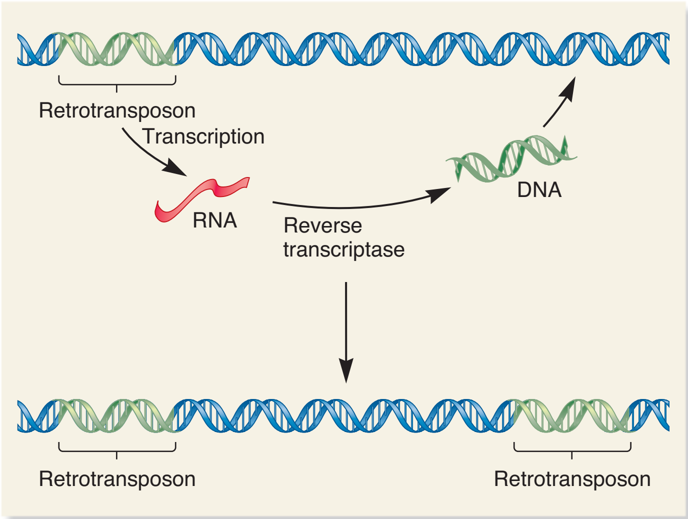
†The numbers of genes were predicted using computer methods described in Chapter 22. These numbers should be considered as estimates of the total gene number.



(a) Simple transposition



(b) Replicative transposition



(c) Retrotransposition

FIGURE 21.16 **Three mechanisms of transposition.**

Concept check: Which of these mechanisms causes the TE to increase in number?

plants. Since that time, geneticists have discovered many different types of TEs in organisms as diverse as bacteria, fungi, plants, and animals. The advent of molecular technology has allowed scientists to understand more about the characteristics of TEs that enable them to be mobile. In this section, we will examine the characteristics of TEs and explore the mechanisms that explain how they move. We will also discuss the biological significance of TEs.

Transposable Elements Move Via One of Three Transposition Pathways

Many different TEs have been found in bacteria, fungi, plants, and animals. Three general types of transposition pathways have been identified (Figure 21.16). In **simple transposition,** also called **conservative transposition,** the TE is removed from its original site and transferred to a new target site. Simple transposition is sometimes called cut-and-paste transposition, because the element is cut out of its original site and pasted into a new one. Transposable elements that move via simple transposition are widely found in bacterial and eukaryotic species. By comparison, **replicative transposition** involves the replication of the TE and insertion of the newly made copy into a second site. In this case, one of the TEs remains in its original location, and the other is inserted at another location. Replicative transposition is relatively uncommon and found only in bacterial species. Both of these types of TEs move as DNA molecules from one site to another. Such TEs are called **transposons.**

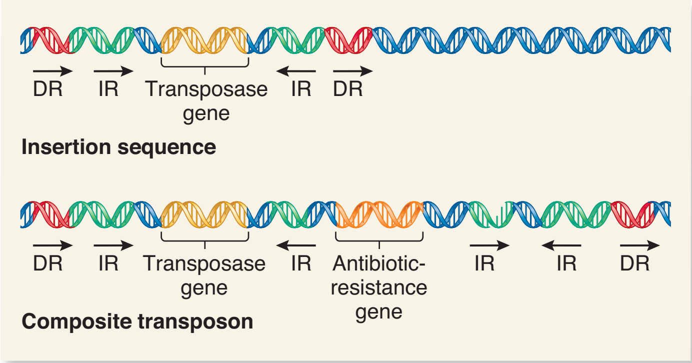
A third category of elements moves via an RNA intermediate. This form of transposition, termed **retrotransposition,** is found in eukaryotic species, where it is very common. Transposable elements that move via retrotransposition are known as **retrotransposons, retroelements,** or **retroposons.** In retrotransposition, the element is transcribed into RNA. An enzyme called reverse transcriptase uses the RNA as a template to synthesize a DNA molecule that is integrated into a new region of the genome. Like TEs that move by replicative transposition, retrotransposons increase in number during retrotransposition.

Each Type of Transposable Element Has a Characteristic Pattern of DNA Sequences

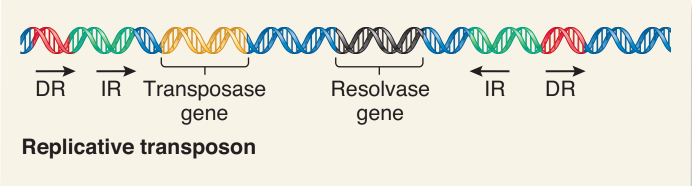
Research on TEs from many species has established that DNA sequences within them are organized in several different ways. Figure 21.17 describes a few of those ways, although many variations are possible. All TEs are flanked by **direct repeats (DRs),** also called target-site duplications, which are identical nucleotide sequences that are oriented in the same direction and repeated. Direct repeats are adjacent to both ends of the element. The simplest TE, which is commonly found in bacteria, is known as an **insertion sequence.** As shown in Figure 21.17a, an insertion sequence has two important characteristics. First, both ends of the insertion sequence contain **inverted repeats (IRs).** Inverted repeats are DNA sequences that are identical (or very similar) but run in opposite directions, such as the following:

5′–CTGACTCTT–3′ and 5′–AAGAGTCAG–3′
3′–GACTGAGAA–5′ 3′–TTCTCAGTC–5′

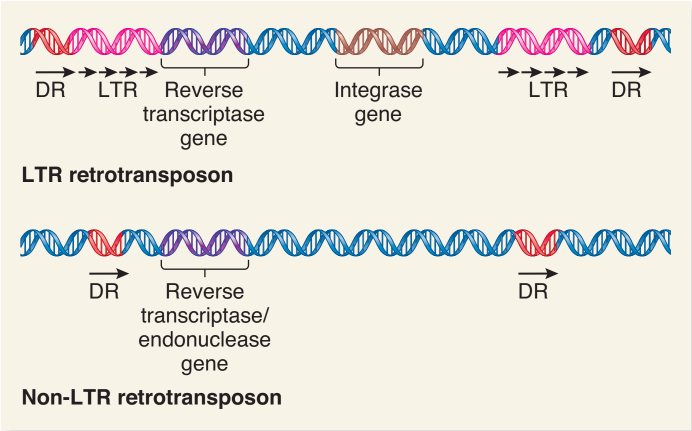
Depending on the particular element, the lengths of inverted repeats range from 9 to 40 bp. In addition, insertion sequences may contain a central region that encodes the enzyme **transposase,** which catalyzes the transposition event.



(a) TEs that move by simple transposition



(b) A TE that moves by replicative transposition



(c) TEs that move by retrotransposition (via an RNA intermediate)

FIGURE 21.17 **Common organizations of transposable elements (TEs).** Direct repeats (DRs) are found at the ends of a TE. Inverted repeats (IRs) are near the ends of many transposable elements. Long terminal repeats (LTRs) are regions containing a large number of tandem repeats.

Composite transposons contain additional genes that are not necessary for transposition. They commonly contain genes that confer a selective advantage to the organism under certain growth conditions. Composite transposons are prevalent in bacteria, where they often contain genes that provide resistance to antibiotics or toxic heavy metals. For example, the composite transposon shown in Figure 21.17a contains two insertion sequences flanking a gene that confers antibiotic resistance. During transposition of a composite transposon, only the inverted repeats at the ends of the transposon are involved in the transposition event. Both insertion sequences and composite transposons are elements that move via simple transposition. **Replicative transposons,** elements that move by replicative transposition, have a sequence organization that is similar to insertion sequences except that replicative transposons have a resolvase gene that is found between the inverted repeats (Figure 21.17b). Both transposase and resolvase are needed to catalyze the transposition of replicative transposons.

The organization of retrotransposons varies greatly, and they are categorized based on their evolutionary relationship to retroviral sequences. Retroviruses are RNA viruses that make a DNA copy that integrates into the host's genome. The **LTR retrotransposons** are evolutionarily related to known retroviruses. These TEs have retained the ability to move around the genome, though, in most cases, they do not produce mature viral particles. LTR retrotransposons are so named because they contain **long terminal repeats (LTRs)** at both ends of the element (Figure 21.17c). The LTRs are typically a few hundred bp in length. Like their viral counterparts, LTR retrotransposons may encode virally related proteins such as reverse transcriptase and integrase that are needed for the retrotransposition process.

By comparison, **non-LTR retrotransposons** appear less like retroviruses in their sequence. They may contain a gene that encodes a protein that has both reverse transcriptase and endonuclease function (see Figure 21.17c). As discussed later, these functions are needed for retrotransposition. Some non-LTR retrotransposons are evolutionarily derived from normal eukaryotic genes. For example, the *Alu* family of repetitive sequences found in humans is derived from a single ancestral gene known as the *7SL RNA* gene (a component of the complex called signal recognition particle, which targets proteins to the endoplasmic reticulum). This gene sequence has been copied by retrotransposition to achieve the current number of approximately 1,000,000 copies.

Transposable elements are considered to be complete or **autonomous elements** when they contain all of the information necessary for transposition or retrotransposition to take place. However, TEs are often incomplete or nonautonomous. A **nonautonomous element** typically lacks a gene such as transposase or reverse transcriptase that is necessary for transposition. For example, McClintock studied a TE in corn called *Ds*, which is a nonautonomous element that lacks a transposase gene. An element that is similar to *Ds* but contains a functional transposase gene is called the *Ac* locus or *Ac* element, which stands for Activator element. An *Ac* element can provide a transposase gene that enables *Ds* to transpose. Therefore, nonautonomous TEs such as *Ds* can transpose only when the *Ac* element is present at another region in the genome.

Transposase Catalyzes the Excision and Insertion of Transposons

Now that we understand the typical organization of TEs, let's examine the steps of the transposition process. The enzyme transposase catalyzes the removal of a TE from its original site in the chromosome and its subsequent insertion at another location. A general scheme for simple transposition is shown in Figure 21.18. Transposase binds to the inverted repeat sequences at the ends of the TE and brings them close together. The DNA is cleaved at the ends of the TE, excising it from its original site within the chromosome. Transposase carries the TE to a new site and cleaves the target DNA sequence at staggered recognition sites. The TE is then inserted and ligated to the target DNA.

As noted in Figure 21.18, the ligation of the TE into its new site initially leaves short gaps in the target DNA. Notice that the DNA sequences in these gaps are complementary to each other (in this case, ATGCT and TACGA). Therefore, when they are filled in by DNA gap repair synthesis, this produces direct repeats that flank both ends of the TE. As noted previously, these direct repeats are common features found adjacent to all TEs (see Figure 21.17).

Although the transposition process depicted in Figure 21.18 does not directly alter the number of TEs, simple transposition is known to increase their number in genomes, in some cases to fairly high levels. How can this happen? The answer is that transposition often occurs around the time of DNA replication (**Figure 21.19**). After a replication fork has passed a region containing a TE, two TEs are found behind the fork—one in each of the replicated regions. One of these TEs could then transpose from its original location into a region ahead of the replication fork. After the replication fork has passed this second region and DNA replication is completed, two TEs are found in one of the chromosomes and one TE in the other chromosome. In this way, simple transposition can lead to an increase in TEs.

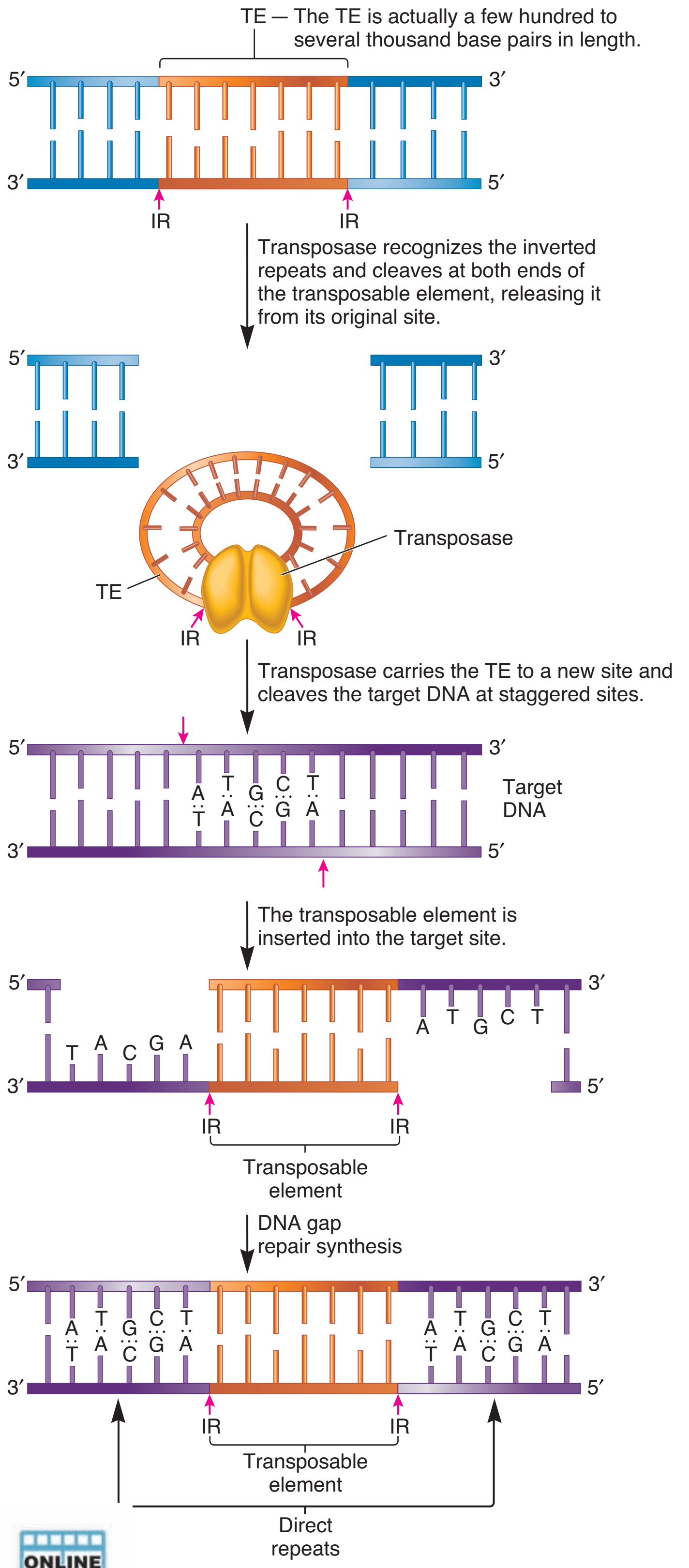

TE — The TE is actually a few hundred to several thousand base pairs in length.
5′
3′
3′
5′
IR
IR
Transposase recognizes the inverted repeats and cleaves at both ends of the transposable element, releasing it from its original site.
5′
3′
3′
5′
Transposase
TE
IR
IR
Transposase carries the TE to a new site and cleaves the target DNA at staggered sites.
5′
3′
A T G C T
T A C G A
Target DNA
3′
5′
The transposable element is inserted into the target site.
5′
3′
T A C G A
A T G C T
3′
5′
IR
IR
Transposable element
DNA gap repair synthesis
5′
3′
A T G C T
T A C G A
A T G C T
T A C G A
3′
5′
IR
IR
Transposable element
Direct repeats

FIGURE 21.18 **Simple transposition.**

Retrotransposons Use Reverse Transcriptase for Retrotransposition

Thus far, we have considered how DNA elements—transposons—can move throughout the genome. By comparison, retrotransposons

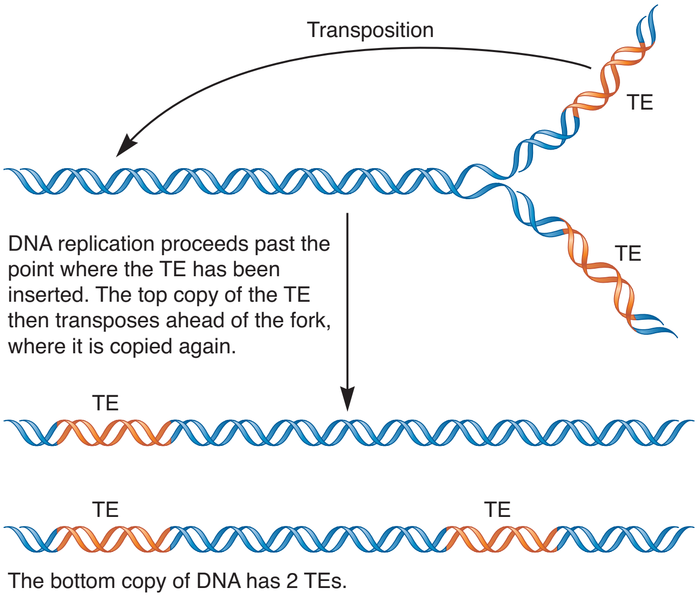

Transposition
TE
TE
DNA replication proceeds past the point where the TE has been inserted. The top copy of the TE then transposes ahead of the fork, where it is copied again.
TE
TE
TE
The bottom copy of DNA has 2 TEs.

FIGURE 21.19 **Increase in transposable element (TE) copy number via simple transposition.** In this example, a TE that has already been replicated transposes to a new site that has not yet replicated. Following the completion of DNA replication, the TE has increased in number.

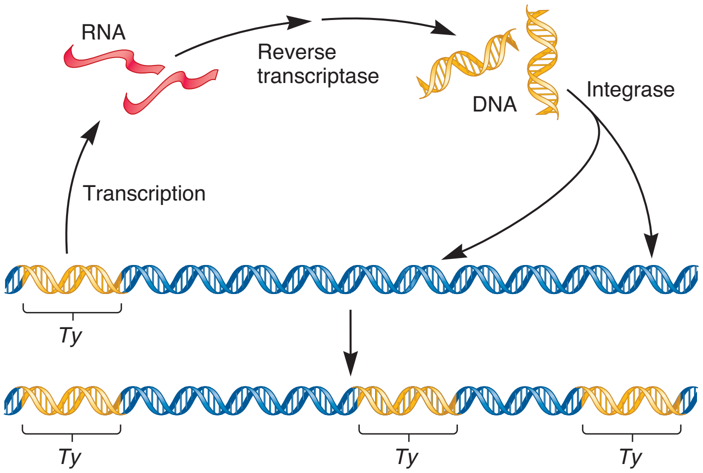


FIGURE 21.20 **Retrotransposition of an LTR retrotransposon.**

Concept check: *What is the function of reverse transcriptase?*

use an RNA intermediate in their transposition mechanism. Let's begin with LTR retrotransposons. As shown in Figure 21.20, the movement of LTR retrotransposons requires two key enzymes: reverse transcriptase and integrase. In this example, the cell already contains a retrotransposon known as *Ty* within its genome. This retrotransposon is transcribed into RNA. In a series of steps, **reverse transcriptase** uses this RNA as a template to synthesize a double-stranded DNA molecule. The LTRs at the ends of the double-stranded DNA are then recognized by **integrase,** which catalyzes the insertion of the DNA into the target chromosomal DNA. The integration of retrotransposons can occur at multiple locations within the genome. (Figure 21.20 shows the integration occurring at two locations.) Furthermore, because a single retrotransposon can be copied into many RNA transcripts, retrotransposons may accumulate rapidly within a genome.

The currently accepted model for the replication and integration of non-LTR retrotransposons is called **target-site primed reverse transcription (TPRT).** As shown in Figure 21.21, the retrotransposon is first transcribed into RNA with a polyA tail. The target DNA site is recognized by an endonuclease, which may be encoded by the retrotransposon. This endonuclease recognizes a consensus sequence of 5′-TTTTA-3′ and initially cuts just one of the DNA strands. The 3′-polyA tail of the retrotransposon RNA can bind to this nicked site due to AT base pairing. Reverse transcriptase then uses the target DNA as a primer and makes a DNA copy of the RNA. This is the process of target-site primed reverse transcription. To be fully integrated into the target DNA, the endonuclease makes a cut in the other DNA strand usually about 7 to 20 nucleotides away from the first cut. The retrotransposon DNA is then ligated into the target site within a chromosome, perhaps by nonhomologous end joining. The mechanism for synthesis of the other DNA strand of the retrotransposon is not completely understood. It may occur via DNA repair mechanisms described in Chapter 18.

Transposable Elements May Have Important Influences on Mutation and Evolution

Over the past few decades, researchers have found that TEs probably occur in the genomes of all species. Table 21.4 describes a few TEs that have been studied in great detail. As discussed in Chapter 12, the genomes of eukaryotic species typically contain moderately and highly repetitive sequences. In some cases, these repetitive sequences are due to the proliferation of TEs. In mammals, for example, **long interspersed elements (LINEs)** are usually 1000 to 10,000 bp in length and found in a few thousand to several hundred thousands of copies. In humans, a particular family of related lines called LINE-1 or L1 is found in about 500,000 copies and represents about 21% of the total human DNA! By comparison, **short interspersed elements (SINEs)** are less than 500 bp in length. A specific example of a SINE is the *Alu* sequence, present in about 1 million copies in the human genome. About 10% of the human genome is composed of this particular TE.

LINEs and SINEs continue to proliferate in the human genome but at a fairly low rate. In about 1 live birth in 100, an *Alu* or L1 sequence has been inserted into a new site in the human genome. On rare occasions, a new insertion can disrupt a gene and cause phenotypic abnormalities. For example, new insertions of *Alu* or L1 sequences into particular genes have been shown, on occasion, to be associated with diseases such as hemophilia, muscular dystrophy, and breast and colon cancer.

The relative abundance of TEs varies widely among different species. As shown in Table 21.5, TEs can be quite prevalent in amphibians, mammals, and flowering plants, but tend to be less abundant in simpler organisms such as bacteria and yeast. The biological significance of TEs in the evolution of prokaryotic and eukaryotic species remains a matter of debate. According to the **selfish DNA hypothesis,** TEs exist because they have characteristics that allow them to multiply within the chromosomal DNA of living cells. In other words, they resemble parasites in the sense that they inhabit a cell without offering any selective advantage to the organism. They can proliferate as long as they do not harm the organism to the extent that they significantly disrupt survival.

Alternatively, other geneticists have argued that most transposition events are deleterious. Therefore, TEs would be eliminated from the genome by natural selection if they did not also offer a compensating advantage. Several potential advantages have been suggested. For example, TEs may cause greater genetic variability by promoting recombination. In addition, bacterial TEs often carry an antibiotic-resistance gene that provides the organism with a survival advantage. Researchers have also suggested that transposition may cause the insertion of exons from one gene into another gene, thereby producing a new gene with novel function(s). This phenomenon, called **exon shuffling,** is described in Chapter 25.

This controversy remains unresolved, but it is clear that TEs can rapidly enter the genome of an organism and proliferate quickly. In *Drosophila melanogaster*, for example, a TE known as the P element was probably introduced into this species in the 1950s. Laboratory stocks of *D. melanogaster* collected prior to

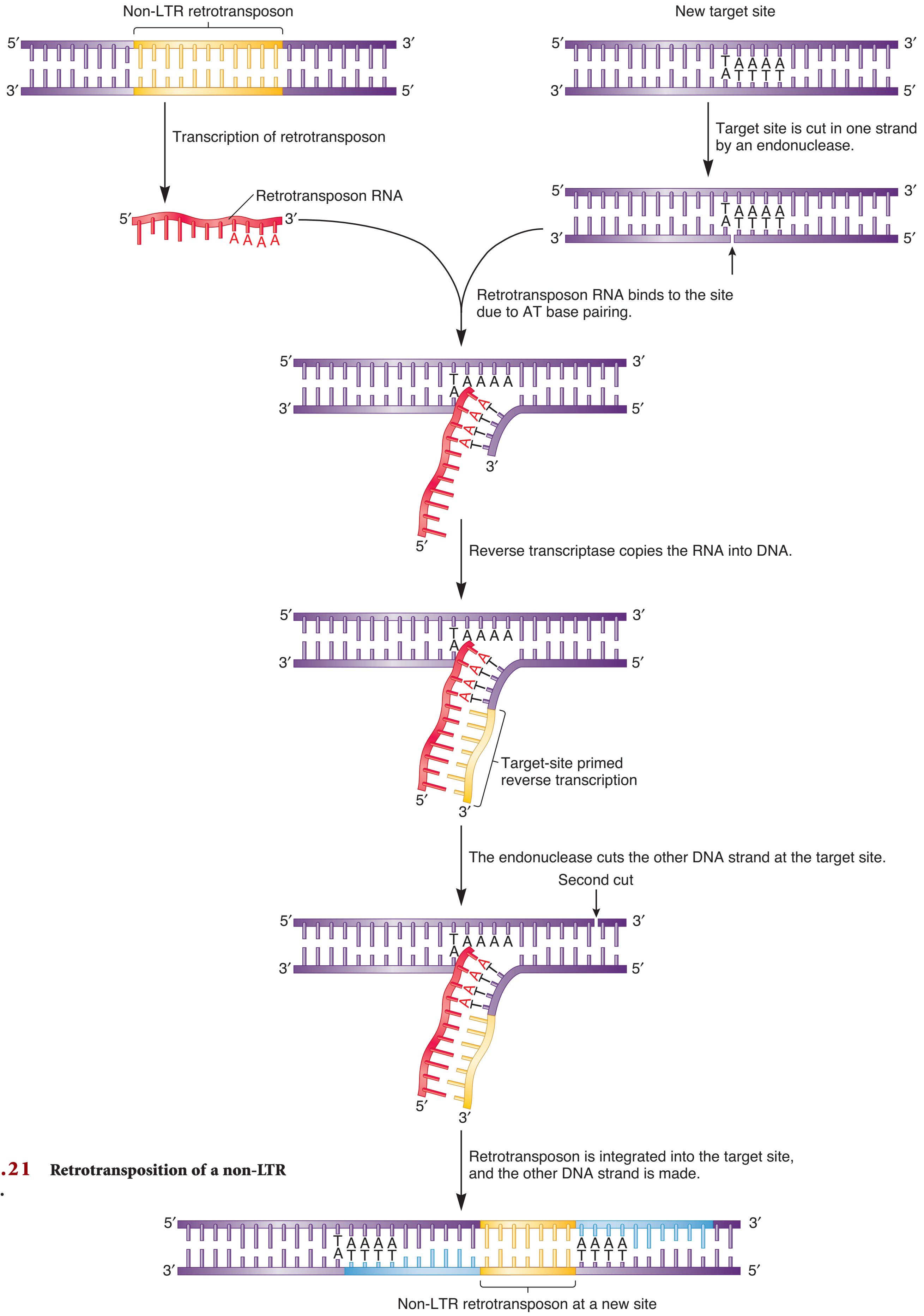

Non-LTR retrotransposon
5′
3′
3′
5′
New target site
TAAAA
ATTTT
Transcription of retrotransposon
Retrotransposon RNA
AAAA
Target site is cut in one strand by an endonuclease.
Retrotransposon RNA binds to the site due to AT base pairing.
Reverse transcriptase copies the RNA into DNA.
Target-site primed reverse transcription
The endonuclease cuts the other DNA strand at the target site.
Second cut
Retrotransposon is integrated into the target site, and the other DNA strand is made.
Non-LTR retrotransposon at a new site

FIGURE 21.21 **Retrotransposition of a non-LTR retrotransposon.**

TABLE 21.4

Examples of Transposable Elements

Element	Type	Approximate Length (bp)	Description
Bacterial			
IS1	Insertion sequence	768	An insertion sequence that is commonly found in five to eight copies in *E. coli*.
Mu	Replicative transposon	36,000	A true virus that can insert itself anywhere in the chromosome of *E. coli* and other bacteria in the Enterobacteriaceae family. Its name, *Mu*, is derived from its ability to insert into genes and mutate them.
Tn10	Composite transposon	9300	One of many different bacterial transposons that carries antibiotic resistance.
Tn951	Composite transposon	16,600	A transposon that provides bacteria with genes that allow them to metabolize lactose.
Yeast			
Ty elements	LTR retrotransposon	6300	A retroelement found in *S. cerevisiae* at about 35 copies per genome.
Drosophila			
P elements	Simple transposon	500–3000	A transposon that may be found in 30–50 copies in P strains of *Drosophila*. It is absent from M strains.
Copia-like elements	LTR retrotransposon	5000–8000	A family of *copia*-like elements found in *Drosophila*, which vary slightly in their lengths and sequences. Typically, each family member is found at about 5–100 copies per genome.
Humans			
Alu sequence	Non-LTR retrotransposon	300	A SINE found in about 1,000,000 copies interspersed throughout the human genome.
L1	Non-LTR retrotransposon	6500	A LINE found in about 500,000 copies interspersed throughout the human genome. Most copies are shorter (truncated) versions.
Plants			
Ac/*Ds*	Simple transposon	4500	*Ac* is an autonomous transposon found in corn and other plant species. It carries a transposase gene. *Ds* is a nonautonomous version that lacks a functional transposase gene.
Opie	LTR retrotransposon	9000	An LTR retrotransposon found in plants that is related to the copia element found in animals.

this time do not contain P elements. Remarkably, in the last 50 years, the P element has expanded throughout *D. melanogaster* populations worldwide. The only strains without the P element are laboratory strains collected prior to the 1950s. This observation underscores the surprising ability of TEs to infiltrate a population of organisms quickly.

Transposable elements have a variety of effects on chromosome structure and gene expression (**Table 21.6**). Because many of these outcomes are likely to be harmful, transposition is usually a highly regulated phenomenon that occurs only in a few individuals under certain conditions. Agents such as radiation, chemical mutagens, and hormones stimulate the movement of TEs. When it is not carefully regulated, transposition is likely to be detrimental. For example, in *D. melanogaster*, if females that lack P elements (M strain females) are crossed with males that contain numerous P elements (P strain males), the egg cells allow the P elements inherited via the sperm to transpose at a high rate. The resulting hybrid offspring exhibit a variety of abnormalities, which include a high rate of sterility, mutation, and chromosome breakage. This deleterious outcome, which is called

TABLE 21.5

Abundance of TEs in the Genomes of Selected Species

Species	Percentage of the Total Genome Composed of Transposable Elements*
Frog (*Xenopus laevis*)	77
Corn (*Zea mays*)	60
Human (*Homo sapiens*)	45
Mouse (*Mus musculus*)	40
Fruit fly (*Drosophila melanogaster*)	20
Nematode (*Caenorhabditis elegans*)	12
Yeast (*Saccharomyces cerevisiae*)	4
Bacterium (*Escherichia coli*)	0.3

*In some cases, the abundance of TEs may vary somewhat among different strains of the same species. The values reported here are typical values.

TABLE 21.6

Possible Consequences of Transposition

Consequence	Cause
Chromosome Structure	
Chromosome breakage	Excision of a TE.
Chromosomal rearrangements	Homologous recombination between TEs located at different positions in the genome.
Gene Expression	
Mutation	Incorrect excision of TEs.
Gene inactivation	Insertion of a TE into a gene.
Alteration in gene regulation	Transposition of a gene next to regulatory sequences or the transposition of regulatory sequences next to a gene.
Alteration in the exon content of a gene	Insertion of exons into the coding sequence of a gene via TEs. This phenomenon is called exon shuffling.
Gene duplications	Creation of a composite transposon that transposes to another site in the genome.

hybrid dysgenesis, occurs because the P elements were able to insert into a variety of locations in the genome.

21.6 REVIEWING THE KEY CONCEPTS

- Transposable elements can move via three general mechanisms called simple transposition (cut-and-paste), replicative transposition, and retrotransposition (see Figure 21.16).
- Each type of transposable element has its own pattern of DNA sequences, which includes direct repeats (see Figure 21.17).
- Transposase catalyzes the excision and insertion of transposons, which move via a DNA intermediate (see Figure 21.18).
- Simple transposition can increase the copy number of a transposon if it occurs just after a transposon has been replicated (see Figure 21.19).
- Retrotransposition of LTR retrotransposons occurs via reverse transcriptase and integrase, whereas retrotransposition of non-LTR retrotransposons occurs via target-site primed reverse transcription (see Figures 21.20, 21.21).
- Many different transposons are found among living organisms. Their abundance varies among different species (see Tables 21.4, 21.5).
- Transposition can have a variety of effects on chromosome structure and gene expression (see Table 21.6).

21.6 COMPREHENSION QUESTIONS

1. Which of the following types of transposable elements involve an RNA intermediate?
   a. Composite transposons
   b. Replicative transposons
   c. Retroelements
   d. All of the above
2. The function of transposase is
   a. to recognize inverted repeats.
   b. to remove a TE from its original site.
   c. to insert a TE into a new site.
   d. all of the above.
3. According to the selfish DNA hypothesis, TEs exist because
   a. they offer the host a selective advantage.
   b. they have characteristics that allow them to multiply within the chromosomal DNA of living cells.
   c. they promote the expression of certain beneficial genes.
   d. all of the above.

KEY TERMS

Page 499. genome, genomics
Page 500. functional genomics, proteomics, mapping, cytogenetic mapping, linkage mapping, physical mapping, genetic map, chromosome map, locus, loci
Page 501. in situ hybridization, fluorescence in situ hybridization (FISH)
Page 502. chromosome painting
Page 503. molecular marker, polymorphic, restriction fragment length polymorphisms (RFLPs)
Page 504. monomorphic
Page 505. lod score method, microsatellites, sequence-tagged site (STS)
Page 506. founder
Page 507. contig
Page 508. artificial chromosomes, yeast artificial chromosome (YAC), bacterial artificial chromosomes (BACs), P1 artificial chromosomes (PACs)
Page 509. cosmid
Page 510. positional cloning
Page 511. chromosome walking, subcloning
Page 512. genome-sequencing projects, shotgun sequencing, hierarchical shotgun sequencing
Page 513. whole-genome shotgun sequencing, double-barrel shotgun sequencing
Page 514. Human Genome Project, high-throughput sequencing
Page 515. next-generation sequencing technologies, pyrosequencing, sequencing by synthesis (SBS)
Page 517. comparative genomics, transposition, transposable elements (TEs)
Page 519. simple transposition, conservative transposition, replicative transposition, transposons, retrotransposition, retrotransposons, retroelements, retroposons, direct repeats (DRs), insertion sequence, inverted repeats (IRs), transposase
Page 520. composite transposons, replicative transposons, LTR retrotransposons, long terminal repeats (LTRs), non-LTR retrotransposons, autonomous elements, nonautonomous element
Page 522. reverse transcriptase, integrase, target-site primed reverse transcription (TPRT), long interspersed elements (LINEs), short interspersed elements (SINEs), selfish DNA hypothesis, exon shuffling
Page 525. hybrid dysgenesis

CHAPTER SUMMARY

- The genome is the total genetic composition of an organism or species. Genomics refers to the molecular analysis of a genome as a whole unit. Functional genomics is aimed at studying the expression of genes, whereas proteomics is focused on the structures and functions of proteins.

21.1 Overview of Chromosome Mapping

- The term *mapping* refers to the process of determining the relative locations of genes or other DNA segments along a chromosome, thereby producing a genetic or chromosome map. The process may involve cytogenetic, linkage, or physical mapping techniques (see Figure 21.1).

21.2 Cytogenetic Mapping Via Microscopy

- Cytogenetic mapping attempts to determine the locations of particular genes relative to a banding pattern of a chromosome.
- Fluorescence in situ hybridization (FISH) is a commonly used method to map genes and other segments along a chromosome (see Figures 21.2, 21.3).

21.3 Linkage Mapping Via Crosses

- Linkage mapping often relies on molecular markers to map the locations of genes along chromosomes (see Table 21.1).
- Restriction fragment length polymorphisms (RFLPs) and microsatellites are used as molecular markers (see Figures 21.4–21.6).

21.4 Physical Mapping Via Cloning

- A physical map of a chromosome is made by producing a contiguous collection of clones called a contig, which spans an entire chromosome (see Figures 21.7, 21.8).
- YACs, BACs, PACs, and cosmids are vectors that can carry a long segment of DNA and are used to make contigs (see Figures 21.9–21.11).
- Chromosome walking is a technique for cloning a gene that is close to a known marker (see Figure 21.12).

21.5 Genome-Sequencing Projects

- Shotgun DNA sequencing has been commonly used to sequence the genomes of many species. Overlapping regions allow researchers to determine the contiguous DNA sequence. Shotgun sequencing can follow a hierarchical or whole-genome shotgun-sequencing approach (see Figures 21.13, 21.14).
- The Human Genome Project resulted in the successful sequencing of the entire human genome.
- Newer methods of DNA sequencing can process many samples of DNA simultaneously and are superseding the dideoxy method. An example is pyrosequencing (see Table 21.2, Figure 21.15).
- Since 1995, the genomes of hundreds of species have been sequenced (see Table 21.3).

21.6 Transposition

- Transposable elements can move via three general mechanisms called simple transposition (cut-and-paste), replicative transposition, and retrotransposition (see Figure 21.16).
- Each type of transposable element has its own pattern of DNA sequences, which includes direct repeats (see Figure 21.17).
- Transposase catalyzes the excision and insertion of transposons, which move via a DNA intermediate (see Figure 21.18).
- Simple transposition can increase the copy number of a transposon if it occurs just after a transposon has been replicated (see Figure 21.19).
- Retrotransposition of LTR retrotransposons occurs via reverse transcriptase and integrase, whereas retrotransposition of non-LTR retrotransposons occurs via target-site primed reverse transcription (see Figures 21.20, 21.21).
- Many different transposons are found among living organisms. Their abundance varies among different species (see Tables 21.4, 21.5).
- Transposition can have a variety of effects on chromosome structure and gene expression (see Table 21.6).

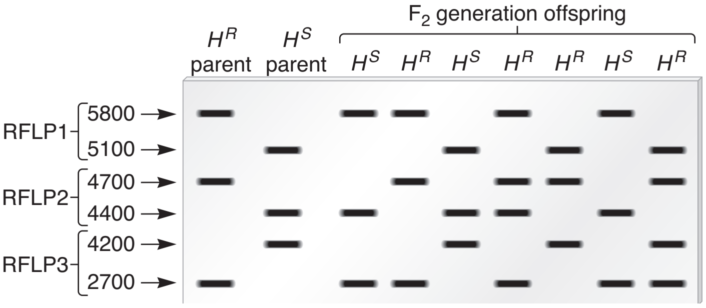
PROBLEM SETS & INSIGHTS

Solved Problems

S1. An RFLP marker is located 1 million bp away from a gene of interest. Your goal is to start at this RFLP marker and walk to this gene. The average insert size in the library is 55,000 bp, and the average overlap at each end is 5,000 bp. Approximately how many steps will it take to get there?

Answer: Each step is only 50,000 bp (i.e., 55,000 minus 5000), because you have to subtract the overlap between adjacent fragments, which is 5000 bp, from the average inset size. Therefore, it will take about 20 steps to go 1 million bp.

S2. When many RFLPs have been mapped within the genome of a plant, an RFLP analysis can be used to map an herbicide-resistance gene. For example, let's suppose that an agricultural geneticist has two strains: one that is herbicide-resistant and the other that is herbicide-sensitive. The two strains differ with regard to many RFLPs. The sensitive and resistant strains are crossed, and the F_1 offspring are allowed to self-fertilize. The F_2 offspring are then analyzed with regard to their herbicide sensitivity and RFLP markers. The following results are obtained:



To which RFLP might the herbicide-resistance gene be linked?

Answer: The aim of this experiment is to correlate the presence of a particular RFLP with the herbicide-resistant phenotype. As shown in the data, the herbicide-resistant parent and all of the herbicide-resistant offspring have an RFLP that is 4700 bp in length. In an actual experiment, a thorough lod analysis would be conducted to determine if linkage is considered likely. If so, the 4700-bp RFLP may either contain the gene that confers herbicide resistance, or, as is more likely, the two may be linked. If the 4700-bp RFLP has already been mapped to a particular site in the plant's genome, the herbicide-resistance gene also maps to the same site or very close to it. This information may then be used by a plant breeder when making future crosses to produce herbicide-resistant plant strains.

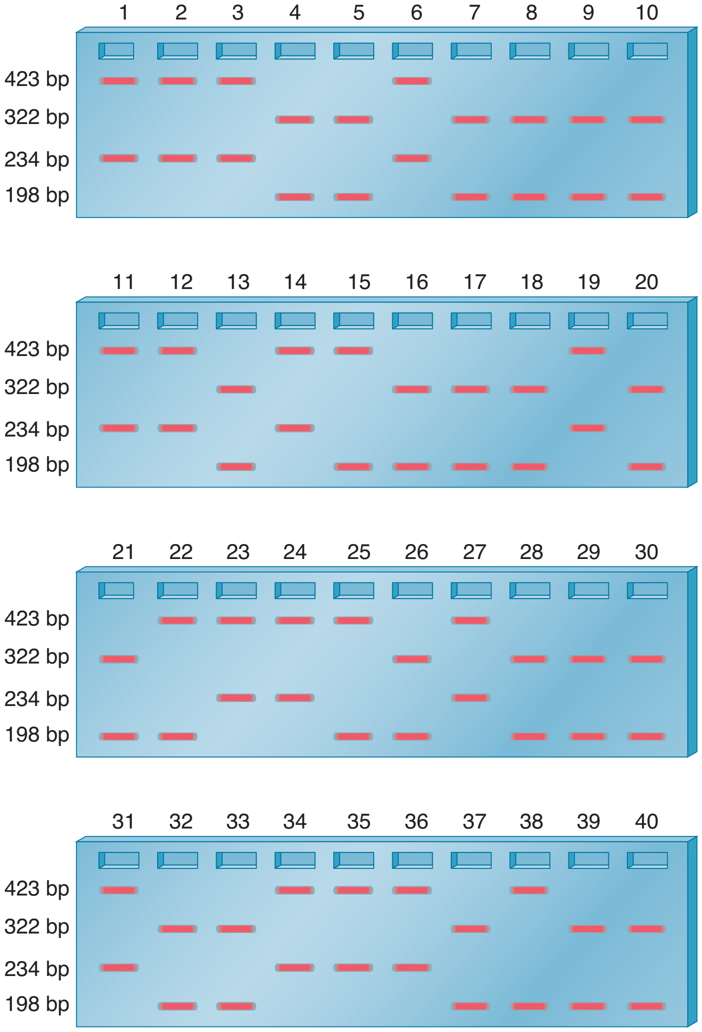
S3. Does a molecular marker have to be polymorphic to be useful in mapping studies? Does a molecular marker have to be polymorphic to be useful in linkage mapping (i.e., involving family pedigree studies or genetic crosses)? Explain why or why not.

Answer: A molecular marker does not have to be polymorphic to be useful in mapping studies. Many sequence-tagged sites (STSs) that are used in physical or cytogenetic studies are monomorphic. Monomorphic markers can provide landmarks in mapping studies.

In linkage mapping studies, a marker must be polymorphic to be useful. Polymorphic molecular markers can be RFLPs, microsatellites, or SNPs. To compute map distances in linkage analysis, individuals must be heterozygous for two or more markers (or genes). For experimental organisms, heterozygotes are testcrossed to homozygotes, and then the number of recombinant offspring and nonrecombinant offspring is determined. For markers that do not assort independently (i.e., linked markers), the map distance is computed as the number of recombinant offspring divided by the total number of offspring times 100.

S4. The distance between two molecular markers that are linked along the same chromosome can be determined by analyzing the outcome of crosses. This can be done in humans by analyzing the members of a pedigree. However, the accuracy of linkage mapping in human pedigrees is fairly limited because the number of people in most families is relatively small. As an alternative, researchers can analyze a population of sperm, produced from a single male, and compute linkage distance in this manner.

As an example, let's suppose a male is heterozygous for two polymorphic STSs. STS-1 exists in two sizes: 234 bp and 198 bp. STS-2 also exists in two sizes: 423 bp and 322 bp. A sample of sperm was collected from this man, and individual sperm were placed into 40 separate tubes. In other words, there was one sperm in each tube. PCR is sensitive enough to allow analysis of DNA in a single sperm! Into each of the 40 tubes were added the primers that amplify STS-1 and STS-2, and then the samples were subjected to PCR. The following results were obtained.



A. What is the arrangement of these two STSs in this individual?

B. What is the map distance between STS-1 and STS-2?

Answer: Keep in mind that mature sperm are haploid, so they have only one copy of STS-1 and one copy of STS-2.

A. If we look at the 40 lanes, most of the lanes (i.e., 36 of them) have either the 234-bp and 423-bp STSs or the 198-bp and 322-bp STSs. This is the arrangement of STSs in this male. One chromosome has STS-1 that is 234 bp and STS-2 that is 423 bp, and the homologous chromosome has STS-1 that is 198 bp and STS-2 that is 322 bp.

B. There are four recombinant sperm, shown in lanes 15, 22, 25, and 38.

$$\text{Map distance} = \frac{4}{40} \times 100$$
$$= 10.0 \text{ mu}$$

Note: This is a relatively easy experiment compared with a pedigree analysis, which would involve contacting lots of relatives and collecting samples from each of them.

S5. The technique of chromosome walking involves the stepwise cloning of adjacent DNA fragments until the gene of interest is reached. This method was used to clone the *CFTR* gene that encodes the cystic fibrosis transmembrane regulator. When mutant, this causes cystic fibrosis (CF), an autosomal recessive disorder in humans. Linkage mapping studies first indicated that the *CFTR* gene was located between two markers designated MET and D7S8. Later mapping studies put the *CFTR* gene between IRP and D7S8. Walking was initiated from the D7S122 marker.

MET D7S122 IRP *CFTR* gene D7S8

(From J. M. Rommens, M. C. Iannuzzi, B. Kerem, et al. (1989) Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science 245,* 1059–65.)

In this study, the researchers restriction-mapped this region using two enzymes, *Not*I and *Xho*I. The locations of sites are shown here:

MET D7S122 IRP *CFTR* gene D7S8

↑A↑ B ↑C↑ D ↑E ↑F ↑

N X X X X X N

N = *Not*I, X = *Xho*I

How would the researchers know they were walking toward the *CFTR* gene and not away from it?

Answer: The general answer is that they had to determine if they were walking toward the IRP marker or toward the MET marker.

As depicted in the previous diagram, digestion of the DNA with *Not*I would produce a large DNA fragment (including regions A–F) that contained both the D7S122 and IRP markers. Because the D7S122 and IRP markers were available as probes, they could confirm that this region contained both markers by Southern blotting of this *Not*I fragment to both markers. *Xho*I cuts this region into smaller pieces. The piece labeled B carried the D7S122 marker, and the piece labeled D carried the IRP marker. Again, this could be confirmed by Southern blotting. This map enabled the researchers to determine in which direction to walk. They began at fragment B (which contained the D7S122 marker) and walked toward the adjacent fragment labeled C. They then walked to the fragment labeled D, which contains the IRP marker. If they had walked toward the fragment labeled A, they would have walked to the MET marker. This is the incorrect direction, because they knew from their linkage mapping that IRP is closer to the *CFTR* gene than to MET.

Conceptual Questions

C1. A person with a rare genetic disease has a sample of her chromosomes subjected to in situ hybridization using a probe that is known to recognize band p11 on chromosome 7. Even though her chromosomes look cytologically normal, the probe does not bind to her chromosomes. How would you explain these results? How would you use this information to positionally clone the gene that is related to this disease?

C2. For each of the following, decide if it could be appropriately described as a genome:

A. the *E. coli* chromosome

B. human chromosome 11

C. a complete set of 10 chromosomes in corn

D. a copy of the single-stranded RNA packaged into human immunodeficiency virus (HIV)

C3. Which of the following statements are true about molecular markers?

A. All molecular markers are segments of DNA that carry specific genes.

B. A molecular marker is a segment of DNA that is found at a specific location in a genome.

C. We can follow the transmission of a molecular marker by analyzing the phenotype (i.e., the individual's bodily characteristics) of offspring.

D. We can follow the transmission of molecular markers using molecular techniques such as gel electrophoresis.

E. An STS is a molecular marker.

C4. If you were examining a sequence of chromosomal DNA, what characteristics would cause you to believe that the DNA contained a transposable element?

C5. Why does transposition produce direct repeats in the chromosomal DNA?

C6. Which types of TEs have the greatest potential for proliferation: insertion sequences, replicative TEs, or retrotransposons? Explain your choice.

C7. Do you consider TEs to be mutagens? Explain.

C8. Let's suppose that a species of mosquito has two different types of simple transposons that we will call X elements and Z elements. The X elements appear quite stable. When analyzing a population of 100 mosquitoes, every mosquito has six X elements, and they are always located in the same chromosomal locations among different individuals. In contrast, the Z elements seem to "move around" quite a bit. Within the same 100 mosquitoes, the number of Z elements ranges from 2 to 14, and the locations of the Z elements tend to vary considerably among different individuals. Explain how one simple transposon can be stable and another simple transposon can be mobile within the same group of individuals.

C9. This chapter describes five different types of TEs: insertion sequences, composite transposons, replicative transposons, LTR retrotransposons, and non-LTR retrotransposons. Which of these five types of TEs would have the following features?

A. Require reverse transcriptase to transpose

B. Require transposase to transpose

C. Are flanked by direct repeats

D. Have inverted repeats

C10. What features distinguish a transposon from a retrotransposon? How are their sequences different, and how are their mechanisms of transposition different?

C11. What is the difference between an autonomous and a nonautonomous TE? Is it possible for nonautonomous TEs to move? If yes, explain how.

C12. The occurrence of multiple transposons within the genome of organisms has been suggested as a possible cause of chromosomal rearrangements such as deletions, duplications, translocations, and inversions. How could the occurrence of transposons promote these kinds of structural rearrangements?

Application and Experimental Questions

E1. Would the following methods be described as linkage, cytogenetic, or physical mapping?

A. Fluorescence in situ hybridization (FISH)

B. Conducting dihybrid crosses to compute map distances

C. Chromosome walking

D. Examination of polytene chromosomes in *Drosophila*

E. Use of RFLPs in crosses

F. Using BACs and cosmids to construct a contig

E2. In an in situ hybridization experiment, what is the relationship between the base sequence of the probe DNA and the site on the chromosomal DNA where the probe binds?

E3. Describe the technique of in situ hybridization. Explain how it can be used to map genes.

E4. The cells from a malignant tumor were subjected to in situ hybridization using a probe that recognizes a unique sequence on chromosome 14. The probe was detected only once in each of these cells. Explain these results and speculate on their significance with regard to the malignant characteristics of these cells.

E5. Figure 21.2 describes the technique of FISH. Why is it necessary to "fix" the cells (and the chromosomes inside of them) to the slides? What does it mean to fix them? Why is it necessary to denature the chromosomal DNA?

E6. Explain how the use of DNA probes with different fluorescence emission wavelengths can be used in a single FISH experiment to map the locations of two or more genes. This method is called chromosome painting. Explain why this is an appropriate term.

E7. A researcher is interested in a gene found on human chromosome 21. Describe the expected results of a FISH experiment using a probe that is complementary to this gene. How many spots would you see if the probe was used on a sample from an individual with 46 chromosomes versus an individual with Down syndrome?

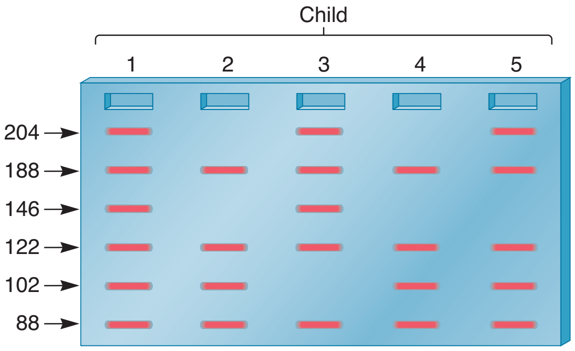
E8. What is a contig? Explain how you would determine that two clones in a contig are overlapping.

E9. Contigs are often made using BAC or cosmid vectors. What are the advantages and disadvantages of these two types of vectors? Which type of contig would you make first, a BAC or cosmid contig? Explain.

E10. Describe the molecular features of a BAC cloning vector. What is the primary advantage of a BAC over a plasmid or viral vector?

E11. In general terms, what is a polymorphism? Explain the molecular basis for a restriction fragment length polymorphism (RFLP). How is an RFLP detected experimentally? Why are RFLPs useful in physical mapping studies? How can they be used to clone a particular gene?

E12. A woman has been married to two different men and produced five children. This group is analyzed with regard to three different STSs: STS-1 is 146 and 122 bp; STS-2 is 102 and 88 bp; and STS-3 is 188 and 204 bp. The mother is homozygous for all three STSs: STS-1 = 122, STS-2 = 88, and STS-3 = 188. Father 1 is homozygous for STS-1 = 122 and STS-2 = 102, and heterozygous for STS-3 = 188/204. Father 2 is heterozygous for STS-1 = 122/146, STS-2 = 88/102, and homozygous for STS-3 = 204. The five children have the following results:



Which children can you definitely assign to father 1 and father 2?

E13. An experimenter used primers to nine different STSs to test their presence along five different BAC clones. The results are shown here.

Alignment of STSs and BACs

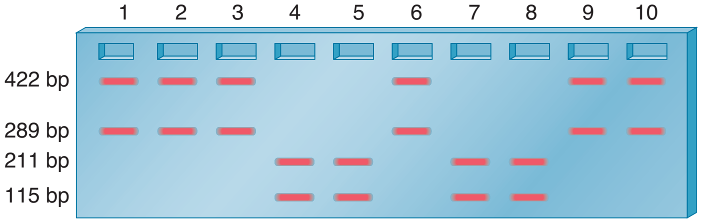
STSs

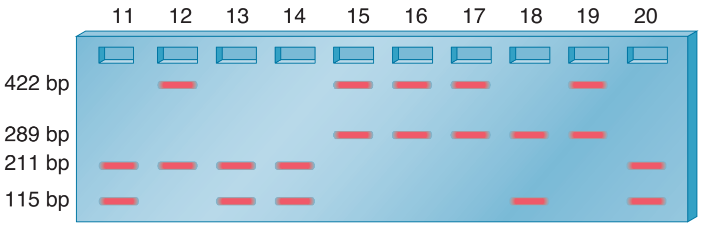
BACs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	–	–	–	–	–	+	+	–	+
2	+	–	–	–	+	–	–	+	–
3	–	–	–	+	–	+	–	–	–
4	–	+	+	–	–	–	–	+	–
5	–	–	+	–	–	–	–	–	+

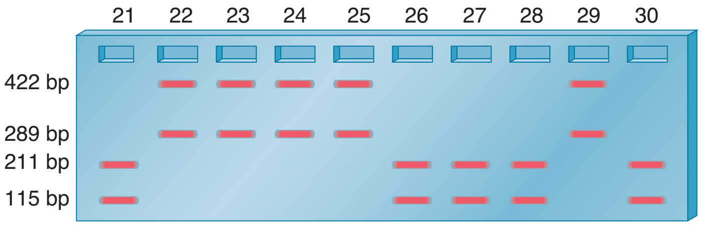
Make a contig map that describes the alignment of the five BACs.

E14. In the Human Genome Project, researchers have collected linkage data from many crosses in which the male was heterozygous for markers and many crosses where the female was heterozygous for markers. The distance between the same two markers, computed in map units or centiMorgans, is different between males and females. In other words, the linkage maps for human males and females are not the same. Propose an explanation for this discrepancy. Do you think the sizes of chromosomes (excluding the Y chromosome) in human males and females are different? How could physical mapping resolve this discrepancy?

E15. Take a look at Solved Problem S4. Let's suppose a male is heterozygous for two polymorphic sequence-tagged sites. STS-1 exists in two sizes: 211 bp and 289 bp. STS-2 also exists in two sizes: 115 bp and 422 bp. A sample of sperm was collected from this man, and individual sperm were placed into 30 separate tubes. Into each of the 30 tubes were added the primers that amplify STS-1 and STS-2, and then the samples were subjected to PCR. The following results were obtained:







A. What is the arrangement of these STSs in this individual?

B. What is the linkage distance between STS-1 and STS-2?

C. Could this approach of analyzing a population of sperm be applied to RFLPs?

E16. Compared with a conventional plasmid, what additional sequences are required in a YAC vector so it can behave like an artificial chromosome? Describe the importance of each required sequence.

E17. When conducting physical mapping studies, place the following methods in their most logical order:

A. Clone large fragments of DNA to make a BAC library.

B. Determine the DNA sequence of subclones from a cosmid library.

C. Subclone BAC fragments to make a cosmid library.

D. Subclone cosmid fragments for DNA sequencing.

E18. Four cosmid clones, which we will call cosmid A, B, C, and D, were subjected to a Southern blot in pairwise combinations. The insert size of each cosmid was also analyzed. The following results were obtained:

Cosmid	Insert Size (bp)	Hybridized to?
A	6000	C
B	2200	C, D
C	11,500	A, B, D
D	7000	B, C

Draw a map that shows the order of the inserts within these four cosmids.

E19. What is an STS? How are STSs generated experimentally? What are the uses of STSs? Explain how a microsatellite can be a polymorphic STS.

E20. A human gene, which we will call gene *X*, is located on chromosome 11 and is found as a normal allele and a recessive disease-causing allele. The location of gene *X* has been approximated on the map shown here that contains four STSs, labeled STS-1, STS-2, STS-3, and STS-4.

STS-1 STS-2 STS-3 Gene *X* STS-4

A. Explain the general strategy of positional cloning.

B. If you applied the approach of positional cloning to clone gene *X*, where would you begin? As you progressed in your cloning efforts, how would you know if you were walking toward or away from gene *X*?

C. How would you know you had reached gene *X*? (Keep in mind that gene *X* exists as a normal allele and a disease-causing allele.)

E21. Describe how you would clone a gene by positional cloning. Explain how a (previously made) contig would make this task much easier.

E22. Discuss the general differences between hierarchical shotgun sequencing and whole-genome shotgun sequencing.

E23. Discuss the advantages of next-generation sequencing technologies.

E24. What is meant by "sequence by synthesis"?

Questions for Student Discussion/Collaboration

1. What is a molecular marker? Give two examples. Discuss why it is easier to locate and map many molecular markers rather than functional genes.
2. Which goals of the Human Genome Project do you think are the most important? Why? Discuss the types of ethical problems that might arise as a result of identifying all of our genes.
3. Based on your current knowledge of genetics, discuss whether or not you think the selfish DNA hypothesis is correct.

Answers to Comprehension Questions

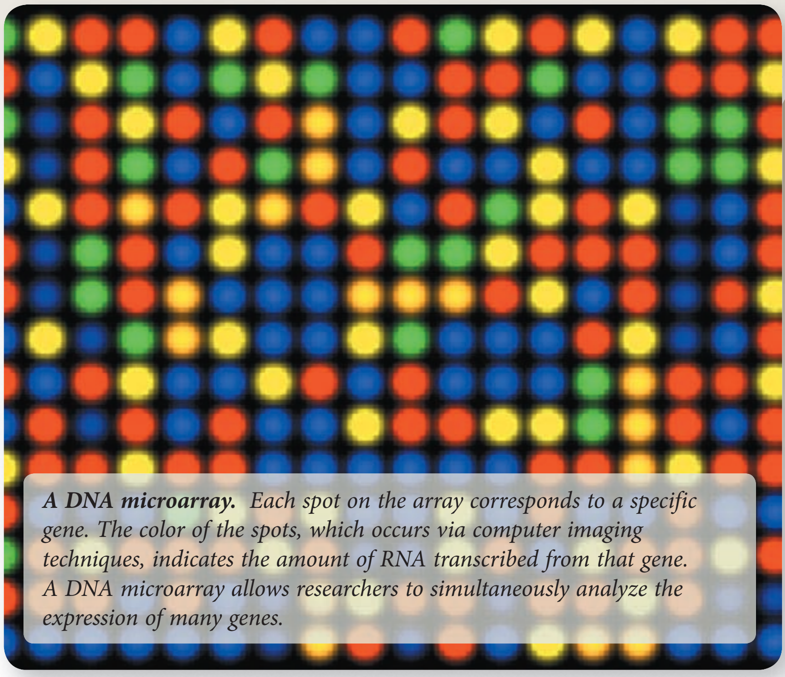
21.1: a
21.2: d
21.3: c, d, a
21.4: b, d, b
21.5: a, b, b
21.6: c, d, b

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

22



A DNA microarray. *Each spot on the array corresponds to a specific gene. The color of the spots, which occurs via computer imaging techniques, indicates the amount of RNA transcribed from that gene. A DNA microarray allows researchers to simultaneously analyze the expression of many genes.*

CHAPTER OUTLINE

22.1 Functional Genomics

22.2 Proteomics

22.3 Bioinformatics

GENOMICS II: FUNCTIONAL GENOMICS, PROTEOMICS, AND BIOINFORMATICS

In Chapter 21, we learned that genomics involves the mapping of an entire genome and, eventually, the determination of a species' complete DNA sequence. The amount of information found within a species' genome is enormous. The goal of **functional genomics** is to elucidate the roles of genetic sequences—DNA, RNA, and amino acid sequences—in a given species. In most cases, functional genomics is aimed at understanding gene function. At the genomic level, researchers can study genes as groups. For example, the information gained from a genome-sequencing project can help researchers study entire metabolic pathways. This provides a description of the ways in which gene products interact to carry out cellular processes. In addition, a study of genetic sequences can help to identify regions that play particular functional roles. For example, an analysis of certain species of bacteria helped to identify DNA sequences that promote the uptake of DNA during bacterial transformation.

Because most genes encode proteins, a goal of many molecular biologists is to understand the functional roles of all of the proteins a species produces. The entire collection of proteins a given cell or organism can make is called its **proteome,** and the study of the function and interactions of these proteins is termed **proteomics.** An objective of researchers in the field of proteomics is to understand the interplay among many proteins as they function to create cells and, ultimately, the traits of a given species.

From a research perspective, functional genomics and proteomics can be broadly categorized in two ways: experimental and computational. The experimental approach involves the study of groups of genes or proteins using molecular techniques in the laboratory. In the first two sections in this chapter, we will focus on these techniques. In the last section, we will consider bioinformatics, which is a computational approach. As a very general definition, **bioinformatics** describes the use of computers, mathematical tools, and statistical techniques to record, store, and analyze biological information. Although we often think of bioinformatics in the context of examining genetic data, such as DNA sequences, bioinformatics can also be applied to information from various sources, such as patient statistics and scientific literature. This rapidly developing branch of biology is highly interdisciplinary, incorporating principles from mathematics, statistics, information science, chemistry, and physics. We will consider the field of bioinformatics and discover how this area has provided great insights into the subjects of functional genomics and proteomics.

22.1 FUNCTIONAL GENOMICS

Though the rapid sequencing of genomes, particularly the human genome, has generated great excitement in the field of genetics, many would argue that an understanding of genome function is fundamentally more interesting. In the past, our ability to study genes involved many of the techniques described in Chapter 19, such as gene cloning, Northern blotting, gel retardation assays, and site-directed mutagenesis. These approaches continue to provide a solid foundation for our understanding of gene function. More recently, genome-sequencing projects have enabled researchers to consider gene function at a more complex level. We now can analyze groups of many genes simultaneously to determine how they work as integrated units that produce the characteristics of cells and the traits of complete organisms. In this section, we examine the use of DNA microarrays, which enable researchers to monitor the expression of thousands of genes simultaneously. We will also discuss why researchers are producing gene knockout collections in which each gene within a given species is inactivated in order to understand gene function at the genomic level.

A Microarray Can Identify Genes That Are Transcribed

Researchers have developed an exciting new technology called a **DNA microarray** (also called a **gene chip**) that makes it possible to monitor the expression of thousands of genes simultaneously. A DNA microarray is a small silica, glass, or plastic slide that is dotted with many different sequences of DNA, each corresponding to a short sequence within a known gene. For example, one spot in a microarray may correspond to a sequence within the β-globin gene, whereas another could correspond to a gene that encodes actin, which is a cytoskeletal protein. A single slide may contain tens of thousands of different spots in an area the size of a postage stamp. The relative location of each spot is known.

How are microarrays made? Some are produced by spotting different samples of DNA onto a slide, much like the way an inkjet printer works. Different DNA fragments, which are made synthetically (e.g., by PCR), are individually spotted onto the slide. The DNA fragments are typically 500 to 5000 nucleotides in length, and a few thousand to tens of thousands are spotted to make a single array. Alternatively, other microarrays contain shorter DNA segments—oligonucleotides—that are directly synthesized on the surface of the slide. In this case, the DNA sequence at a given spot is produced by selectively controlling the growth of the oligonucleotide using narrow beams of light. Such oligonucleotides are typically 25 to 30 nucleotides in length. Hundreds of thousands of different spots can be found on a single array. Overall, the technology of making DNA microarrays is quite amazing.

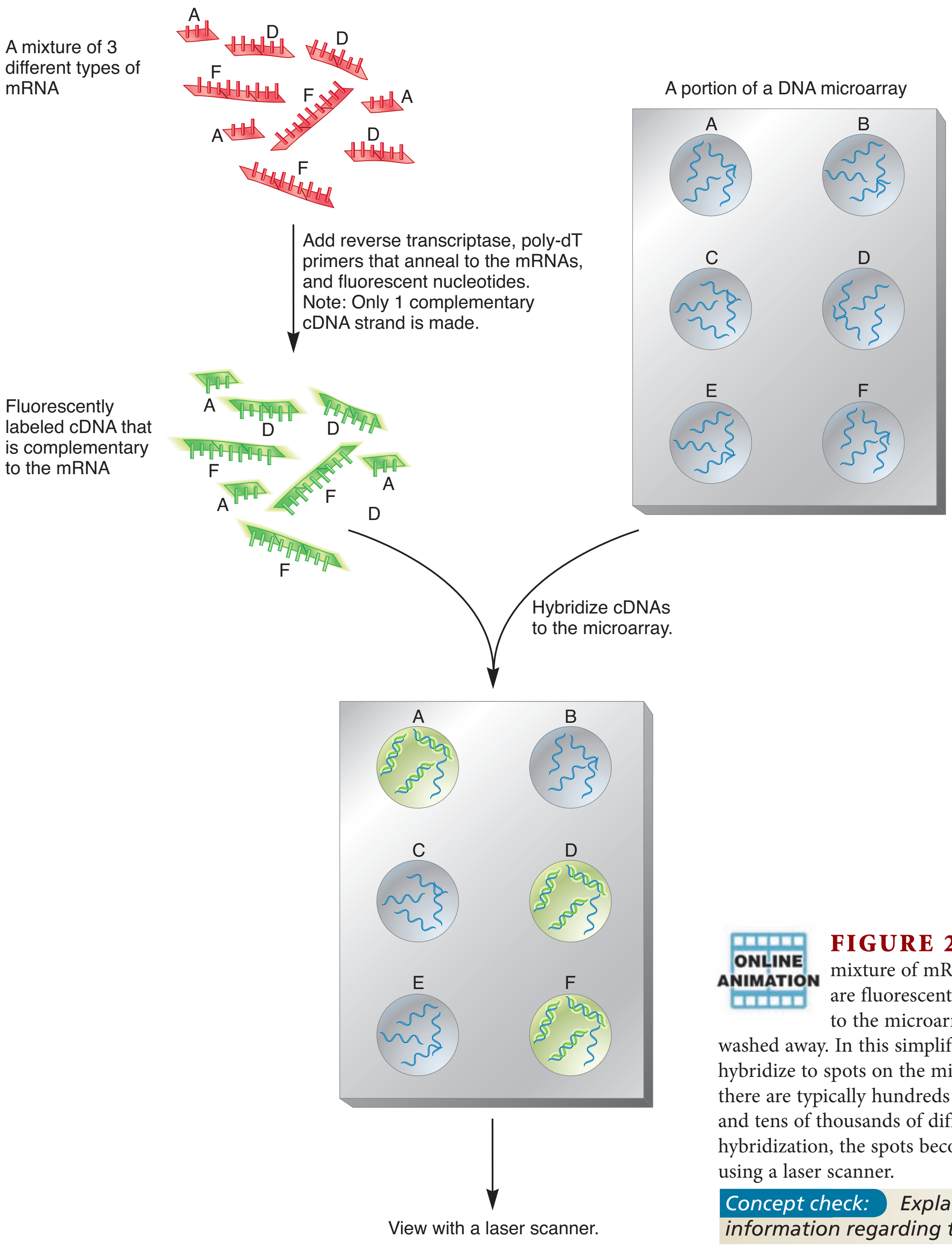
Once a DNA microarray has been made, it is used as a hybridization tool, as shown in **Figure 22.1**. In this experiment, mRNA was isolated from a sample of cells and then used to make fluorescently labeled cDNA. The labeled cDNAs were then layered onto a DNA microarray. Those cDNAs that are complementary to the DNAs in the microarray hybridize, thereby remaining bound to the microarray. The array is then washed with a buffer to remove any unbound cDNAs and placed in a device called a laser scanner, which produces higher resolution images than a conventional optical microscope. The device scans each pixel—the smallest element in a visual image—and after correction for local background, the final fluorescence intensity for each spot is obtained by averaging across the pixels in each spot. This results in a group of fluorescent spots at defined locations in the microarray. High fluorescence intensity in a particular spot means that a large amount of cDNA in the sample hybridized to the DNA at that location. Because the DNA sequence of each spot is already known, a fluorescent spot identifies cDNAs that are complementary to those DNA sequences. Furthermore, because the cDNA was generated from mRNA, this technique identifies RNAs that have been made in a particular cell type under a given set of conditions.

The technology of DNA microarrays has found many important uses (**Table 22.1**). Thus far, its most common use is for studying gene expression patterns. This helps us to understand how genes are regulated in a cell-specific manner and how environmental conditions can induce or repress the transcription of genes. In some cases, microarrays can even help to elucidate the genes encoding proteins that participate in a complicated metabolic pathway. Microarrays can also be used as identification tools. For example, gene expression patterns can aid in the categorization of tumor types. Such identification can be important in the treatment of the disease. Instead of using labeled cDNA, researchers can also hybridize labeled genomic DNA to a microarray. This can be used to identify mutant alleles in a population and to detect deletions and duplications. In addition, this technology is proving useful in correctly identifying closely related bacterial species and subspecies. Finally, microarrays can be used to study DNA-protein interactions as described next.

DNA Microarrays Can Be Used to Identify DNA-Protein Binding at the Genome Level

As discussed throughout this textbook, the binding of proteins to specific DNA sites is critical for a variety of molecular processes, including gene transcription and DNA replication. To study these processes, researchers have devised a variety of techniques for identifying whether or not specific proteins bind to particular sites in the DNA. For example, the techniques of gel retardation assay and DNaseI footprinting, which are described in Chapter 19, are used for this purpose (see Figures 19.14 and 19.15).

More recently, a newer approach called **chromatin immunoprecipitation (ChIP)** has gained widespread use in the analysis of DNA-protein interactions. This method can determine whether proteins bind to a particular region of DNA. A distinguishing feature of this method is that it analyzes DNA-protein


A mixture of 3 different types of mRNA
A
D
D
F
F
A
A
D
F
A portion of a DNA microarray
A
B
C
D
E
F
Add reverse transcriptase, poly-dT primers that anneal to the mRNAs, and fluorescent nucleotides.
Note: Only 1 complementary cDNA strand is made.
Fluorescently labeled cDNA that is complementary to the mRNA
A
D
D
F
A
F
A
D
F
Hybridize cDNAs to the microarray.
A
B
C
D
E
F
View with a laser scanner.

ONLINE ANIMATION

FIGURE 22.1 **A DNA microarray.** A mixture of mRNAs is used to create cDNAs that are fluorescently labeled. The cDNAs are applied to the microarray, and any unbound cDNAs are washed away. In this simplified example, three cDNAs specifically hybridize to spots on the microarray. In an actual experiment, there are typically hundreds or thousands of different cDNAs and tens of thousands of different spots on the array. After hybridization, the spots become fluorescent and can be visualized using a laser scanner.

Concept check: *Explain how this experiment provides information regarding the expression of genes.*

interactions as they occur in the chromatin of living cells. In contrast, a gel retardation assay and DNaseI footprinting are in vitro techniques, which typically use cloned DNA and purified proteins.

Figure 22.2 describes the steps of chromatin immunoprecipitation when it is used in conjunction with microarrays. Proteins in living cells, which are noncovalently bound to DNA, can be more tightly attached to the chromatin by the addition of formaldehyde or some other method that covalently crosslinks the protein to the DNA. Following crosslinking, the cells are lysed, and the DNA is broken by sonication into pieces of approximately 200 to 1000 bp long. Next, an antibody is added that is specific for the protein of interest. To conduct a ChIP assay, a researcher must already suspect that a particular protein binds to the DNA and has previously had an antibody made against that protein. Antibodies are molecules that specifically bind to antigens. In this case, the antigen is the protein of interest that is thought to be a DNA-binding protein. In Figure 22.2, the antibodies are attached to heavy beads. The antibodies bind to

TABLE **22.1**

Applications of DNA Microarrays

Application	Description
Cell-specific gene expression	A comparison of microarray data using cDNAs derived from RNA of different cell types can identify genes that are expressed in a cell-specific manner.
Gene regulation	Environmental conditions play an important role in gene regulation. A comparison of microarray data may reveal genes that are induced under one set of conditions and repressed under another.
Elucidation of metabolic pathways	Genes that encode proteins that participate in a common metabolic pathway are often expressed in a parallel manner. This can be revealed from a microarray analysis. This application overlaps with the study of gene regulation via microarrays.
Tumor profiling	Different types of cancer cells exhibit striking differences in their profiles of gene expression. This can be revealed by a DNA microarray analysis. This approach is gaining widespread use as a method of subclassifying tumors that are sometimes morphologically indistinguishable. Tumor profiling may provide information that can improve a patient's clinical treatment.
Genetic variation	A mutant allele may not hybridize to a spot on a microarray as well as a wild-type allele. Therefore, microarrays are gaining widespread use as a tool for detecting genetic variation. They have been used to identify disease-causing alleles in humans and mutations that contribute to quantitative traits in plants and other species. In addition, microarrays are used to detect chromosomal deletions and duplications.
Microbial strain identification	Microarrays can distinguish between closely related bacterial species and subspecies.
DNA-protein binding	Chromatin immunoprecipitation, which is described in Figure 22.2, can be used with DNA microarrays to determine where in the genome a particular protein binds to the DNA.

DNA-protein complexes and cause the complexes to form a pellet following centrifugation. (See the Appendix for a description of centrifugation.) Because an antibody is made by the immune system of an animal, this step is called immunoprecipitation.

The next step is to identify the DNA to which the protein is covalently crosslinked. To do so, the protein is removed by treatment with chemicals that break the covalent crosslinks. Because the DNA is usually present in very low amounts, researchers must amplify the DNA to analyze it. This is done using PCR (described in Chapter 19). If researchers already suspect that a protein binds to a known

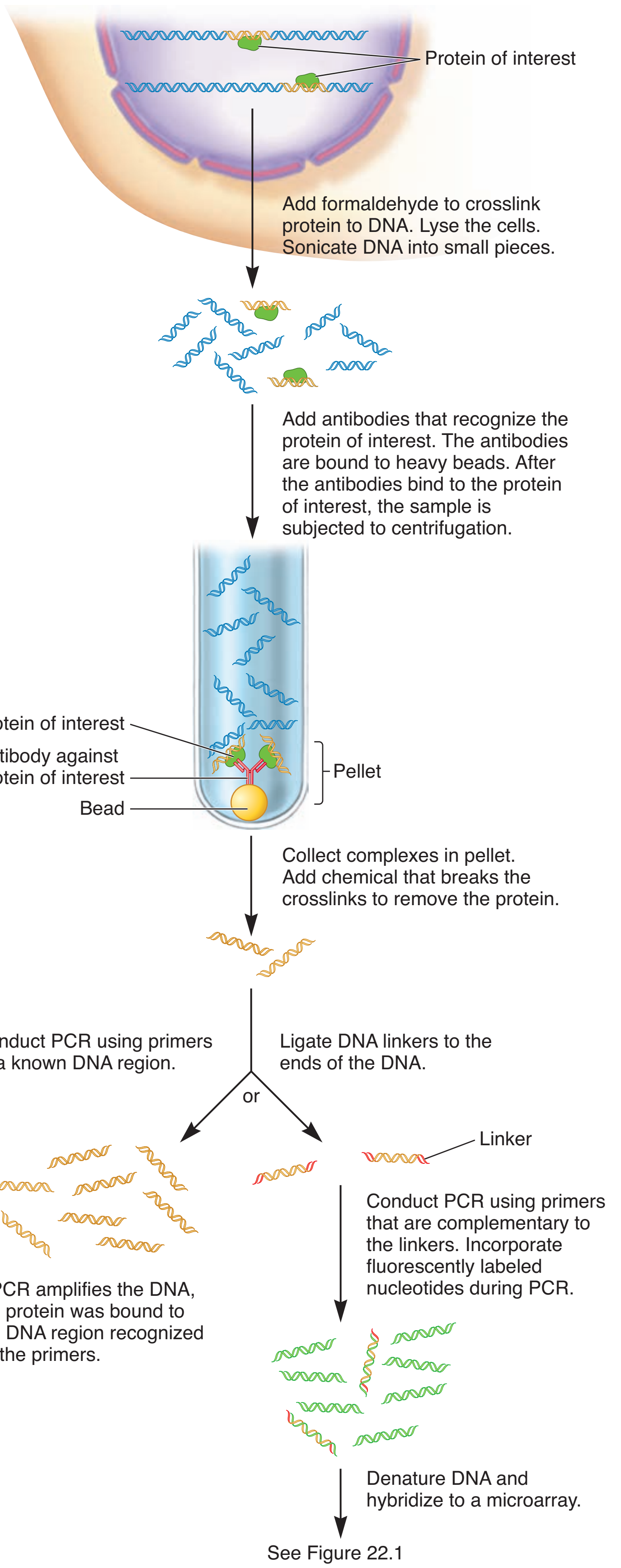

Protein of interest
Add formaldehyde to crosslink protein to DNA. Lyse the cells. Sonicate DNA into small pieces.
Add antibodies that recognize the protein of interest. The antibodies are bound to heavy beads. After the antibodies bind to the protein of interest, the sample is subjected to centrifugation.
Protein of interest
Antibody against protein of interest
Bead
Pellet
Collect complexes in pellet. Add chemical that breaks the crosslinks to remove the protein.
Conduct PCR using primers to a known DNA region.
Ligate DNA linkers to the ends of the DNA.
or
Linker
If PCR amplifies the DNA, the protein was bound to the DNA region recognized by the primers.
Conduct PCR using primers that are complementary to the linkers. Incorporate fluorescently labeled nucleotides during PCR.
Denature DNA and hybridize to a microarray.
See Figure 22.1

FIGURE 22.2 **Chromatin immunoprecipitation (ChIP).** This method can determine whether proteins bind to a particular region of DNA found within the chromatin of living cells. If a microarray is used, as indicated in the bottom right, this is called a ChIP-chip assay.

Concept check: Why is an antibody used in this experiment?

DNA region, they can use PCR primers that specifically flank the DNA region (see bottom left side of Figure 22.2). If they obtain a PCR product, this means that the protein of interest must have been bound (either directly or through other proteins) to this DNA site in living cells.

Alternatively, a researcher may want to determine where the protein of interest binds across the whole genome. In this case, a DNA microarray can be used (see bottom right side of Figure 22.2). Because a DNA microarray is found on a chip, this form of chromatin immunoprecipitation is called a **ChIP-chip assay.** The ends of the precipitated DNA are first ligated to short DNA pieces called linkers. PCR primers are then added that are complementary to the linkers and, therefore, amplify the DNA regions between the linkers. During PCR, the DNA is fluorescently labeled. The labeled DNA is then denatured and hybridized to a DNA microarray. Because the DNA was isolated using an antibody against the protein of interest, the fluorescent spots on the microarray identify sites in the genome where the protein binds. In this way, researchers may be able to determine where a protein binds to locations in the genome, even if those site(s) had not been previously determined by other methods.

Gene Knockout Collections Allow Researchers to Study Gene Function at the Genomic Level

One broad goal of functional genomics is to determine the functions of all the genes in a species' genome. Because each species has thousands of different genes, this is a very complicated task. One approach to achieving this goal is to produce collections of organisms of the same species in which each strain has one of its genes knocked out. For example, in *E. coli*, which has 4377 different genes, a complete knockout collection would be composed of 4377 different strains, with a different gene knocked out in each strain. Recall from Chapter 20 that a **gene knockout** refers to a gene that has been altered in a way that inactivates its function.

Why are knockout collections useful? Consider, for example, a phenotype produced by a particular gene knockout in mice, which cause deafness. Such a result suggests that the function of the normal gene is to promote hearing. Geneticists may also produce knockouts involving two or more genes to understand how the protein products of genes participate in a particular cellular pathway or contribute to a complex trait.

Knockout collections can be made in different ways. One way is via transposable elements, which are described in Chapter 21. When a transposable element hops into a gene, it often inactivates its function. Another way to produce knockouts is via homologous recombination. In Chapter 20, we examined how mice knockouts are made using this approach (refer back to Figure 20.5). In 2006, the National Institutes of Health (NIH) launched the Knockout Mouse Project. The goal of this program is to build a comprehensive and publicly available resource comprising a collection of mouse embryonic stem cells (ES cells) containing a loss-of-function mutation in every gene in the mouse genome. The NIH Knockout Mouse Project will collaborate with other large-scale efforts that are under way in Canada, called the North American Conditional Mouse Mutagenesis Project (NorCOMM), and in Europe, called the European Conditional Mouse Mutagenesis Program (EUCOMM). The collective goal of these programs is to create at least one loss-of-function mutation in each of the approximately 20,000 genes in the mouse genome.

Knockout collections are currently available for several other species, including *E. coli*, *Saccharomyces cerevisiae*, and *Caenorhabditis elegans*. The use of these knockouts has already shed considerable light on the functions of many different genes in these species. Likewise, knockout collections will continue to facilitate our understanding of gene function at the genomic level in the future.

22.1 REVIEWING THE KEY CONCEPTS

- The goal of functional genomics is to understand the role of genetic sequences in a given species.
- A microarray is a slide dotted with many DNA sequences. It can be used to study the expression of many genes simultaneously and also has other uses (see Figure 22.1, Table 22.1).
- In a ChIP-chip assay, chromatin immunoprecipitation is used in conjunction with a microarray to study DNA-protein interactions at the genomic level (see Figure 22.2).
- Researchers are producing gene knockout collections for certain species such as mice to determine the functions of genes at the genomic level.

22.1 COMPREHENSION QUESTIONS

1. A DNA microarray is a slide that is dotted with
   a. mRNA from a sample of cells.
   b. fluorescently labeled cDNA.
   c. known sequences of DNA.
   d. known cellular proteins.
2. To study gene expression from a particular sample of cells, what material is usually hybridized to a DNA microarray?
   a. Fluorescently labeled mRNA
   b. Fluorescently labeled cDNA
   c. Radiolabeled mRNA
   d. Radiolabeled cDNA
3. The purpose of a ChIP-chip assay is to determine
   a. the expression levels of particular genes in the genome.
   b. the amount of a specific protein that is made in a given cell type.
   c. the sites in a genome where a particular protein binds.
   d. all of the above.
4. A gene knockout is a gene
   a. whose function has been inactivated.
   b. that has been transferred to a different species.
   c. that has been moved to a new location in the genome.
   d. that has been eliminated from a species during evolution.

22.2 PROTEOMICS

Thus far, we have considered ways to characterize the genome of a given species and study its function. Because most genes encode proteins, a logical next step is to examine the functional roles of the proteins that a species can make. As mentioned, this field is called proteomics, and the entire collection of a cell's or organism's proteins is its proteome.

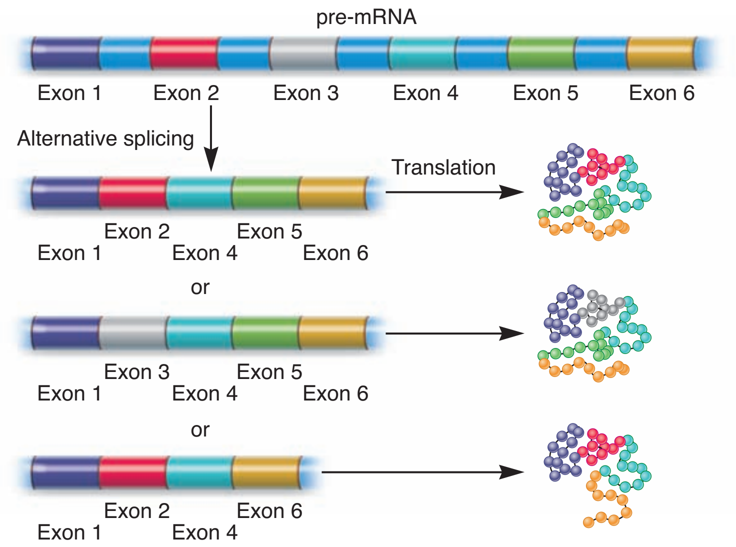
Genomics represents only the first step in our comprehensive understanding of protein structure and function. Researchers often use genomic information to initiate proteomic studies, but such information must be followed up with research that involves the direct analysis of proteins. For example, as discussed in the preceding section, a DNA microarray may provide insights into the transcription of particular genes under a given set of conditions. However, mRNA levels may not provide an accurate measure of the abundance of a protein encoded by a given gene. Protein levels are greatly affected, not only by the level of mRNA, but also by the rate of mRNA translation and by the turnover rate of a given protein. Therefore, DNA microarray data must be corroborated using other methods, such as Western blotting (discussed in Chapter 19), which directly determine the abundance of a protein in a given cell type.

A second way genomic data can elucidate the workings of the proteome is via homology. As discussed later in this chapter, homology between the genes of different species can be used to predict protein structure and/or function when the structure or function of the protein in one of the species is already known. However, homology may not provide direct information regarding the regulation of protein structure and function. Also, it may not reveal potential types of protein-protein interactions in which a given protein may participate. Therefore, even though homology is profoundly useful as a guiding tool, a full understanding of protein function must involve the direct analysis of proteins as they are found in living cells.

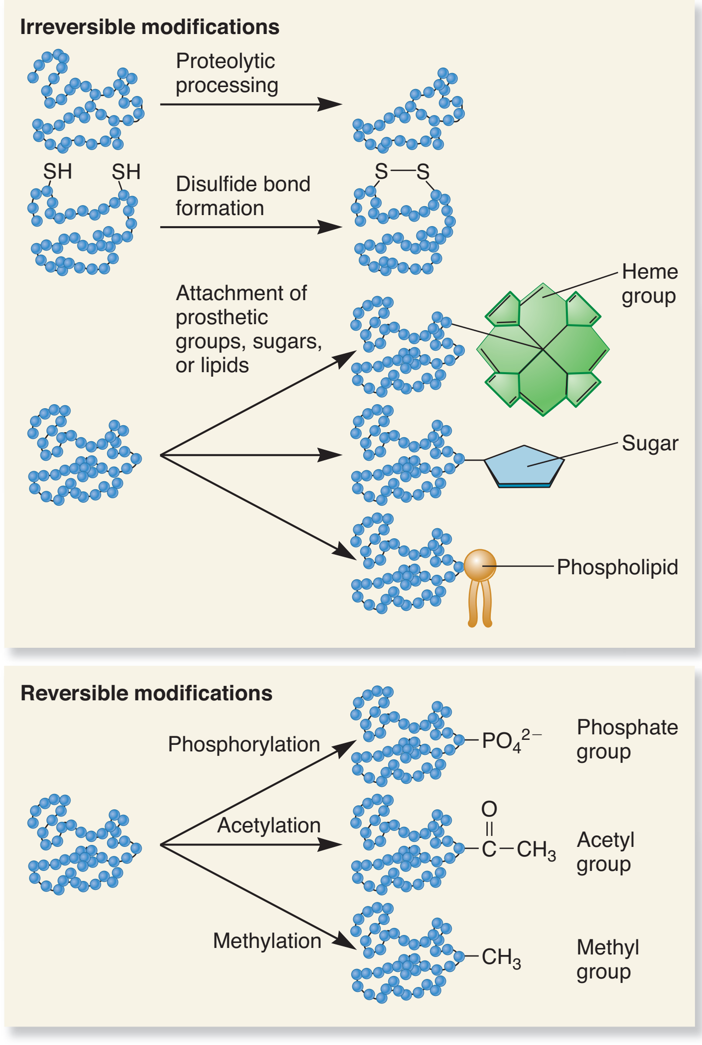
As we move forward in the twenty-first century, the study of proteomes represents a key challenge facing molecular biologists. Much like genomic research, proteomics will require the collective contributions of many research scientists, as well as improvements in technologies that are aimed at unraveling the complexities of the proteome. In this section, we will discuss the phenomena that increase protein diversity beyond genetic diversity. In addition, we will see how the techniques of two-dimensional gel electrophoresis and mass spectrometry can isolate and identify cellular proteins, and explore the use of protein microarrays to study protein expression and function.

The Proteome Is Much Larger Than the Genome

From the sequencing and analysis of an entire genome, researchers can identify all or nearly all of the genes from a given species. The proteome, however, is larger than the genome, and its actual size is somewhat more difficult to determine. What phenomena account for the larger size of the proteome? First, changes in pre-mRNAs may ultimately affect the resulting amino acid sequence of a protein. The most important alteration that occurs commonly in eukaryotic species is **alternative splicing** (**Figure 22.3a**). For many genes, a single pre-mRNA can be spliced into more than one version (also see Chapter 17, Figure 17.17). The splicing is often


pre-mRNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Alternative splicing
Translation
Exon 1
Exon 2
Exon 4
Exon 5
Exon 6
or
Exon 1
Exon 3
Exon 4
Exon 5
Exon 6
or
Exon 1
Exon 2
Exon 4
Exon 6

(a) Alternative splicing


Irreversible modifications
Proteolytic processing
SH
SH
Disulfide bond formation
S—S
Attachment of prosthetic groups, sugars, or lipids
Heme group
Sugar
Phospholipid
Reversible modifications
Phosphorylation
$-PO_4^{2-}$
Phosphate group
Acetylation
$-\overset{\overset{\displaystyle O}{\|}}{C}-CH_3$
Acetyl group
Methylation
$-CH_3$
Methyl group

(b) Posttranslational covalent modification

FIGURE 22.3 **Cellular mechanisms that increase protein diversity. (a)** During alternative splicing, the pattern of exons that remains in a mature mRNA can be different, creating multiple types of transcripts from the same gene. **(b)** In posttranslational covalent modification, after a protein is made, it can be modified in a variety of ways, some of which are irreversible and some reversible.

Concept check: *Explain how these mechanisms affect protein diversity.*

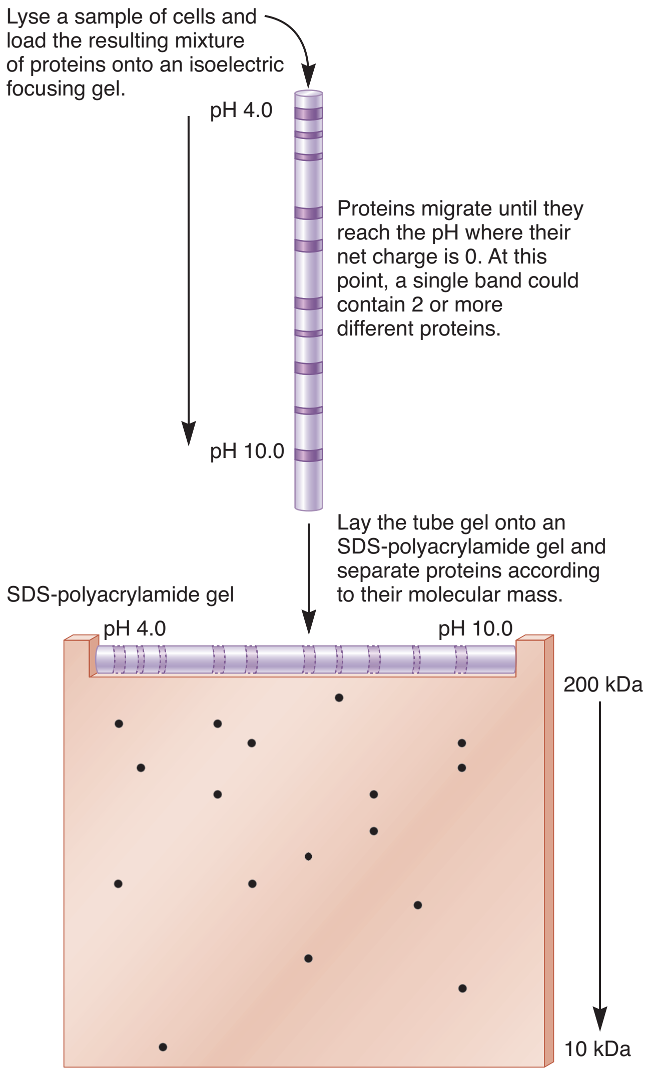
cell-specific, or it may be related to environmental conditions. Alternative splicing is widespread, particularly among complex multicellular organisms. It can lead to the production of several or perhaps dozens of different polypeptide sequences from the same pre-mRNA, which greatly increases the number of potential proteins in the proteome. Similarly, the phenomenon of **RNA editing,** a change in the base sequence of RNA after it has been transcribed (see Chapter 14, Figure 14.23), can lead to changes in the coding sequence of an mRNA. However, RNA editing is much less common than alternative splicing.

Another process that greatly diversifies the composition of a proteome is the phenomenon of **posttranslational covalent modification** (**Figure 22.3b**)**.** Certain types of modifications are involved with the assembly and construction of a functional protein. These alterations include proteolytic processing, disulfide bond formation, and the attachment of prosthetic groups, sugars, or lipids. These are typically irreversible changes that are necessary to produce a functional protein. Other types of modifications, such as phosphorylation, acetylation, and methylation, are often reversible modifications that transiently affect the function of a protein. A protein may be subject to several different types of covalent modification, which can greatly increase the forms of a particular protein found in a cell at any given time.

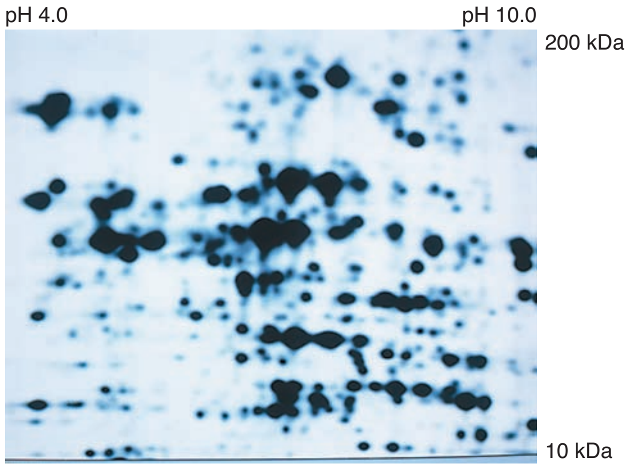
Two-Dimensional Gel Electrophoresis Is Used to Separate a Mixture of Cellular Proteins

As we have just learned, the proteome is usually much larger than a species' genome. Even so, any given cell within a complex multicellular organism produces only a subset of the proteins found in the proteome of a species. For example, the human genome has approximately 20,000 to 25,000 different genes, yet a muscle cell expresses only a subset, and makes about 15,000 types of proteins. The proteins a cell makes depend primarily on the cell type, the stage of development, and the environmental conditions. An objective of researchers in the field of proteomics is the identification and functional characterization of all of the proteins a cell type can make. Because cells produce thousands of different proteins, this is a daunting task. Nevertheless, along with genomic research, the past decade has seen important advances in our ability to isolate and identify cellular proteins.

A common technique in the field of proteomics is **two-dimensional gel electrophoresis,** a method that can separate hundreds or even thousands of different proteins within a cell extract. The steps in this procedure are shown in **Figure 22.4**. As its name suggests, the technique involves two different gel electrophoresis procedures. A sample of cells is lysed, and the proteins are loaded onto the top of a tube gel that separates proteins according to their net charge at a given pH. A protein migrates to the point in the gel where its net charge is zero. This


Lyse a sample of cells and load the resulting mixture of proteins onto an isoelectric focusing gel.
pH 4.0
pH 10.0
Proteins migrate until they reach the pH where their net charge is 0. At this point, a single band could contain 2 or more different proteins.
Lay the tube gel onto an SDS-polyacrylamide gel and separate proteins according to their molecular mass.
SDS-polyacrylamide gel
pH 4.0
pH 10.0
200 kDa
10 kDa

(a) The technique of two-dimensional gel electrophoresis


pH 4.0
pH 10.0
200 kDa
10 kDa

(b) A two-dimensional gel that has been stained to visualize proteins

ONLINE ANIMATION **FIGURE 22.4 Two-dimensional gel electrophoresis.** (**a**) The technique involves two electrophoresis steps. First, a mixture of proteins is separated on an isoelectric-focusing gel that has the shape of a tube. Proteins migrate to the pH in the gel where their net charge is zero. This tube gel is placed into a long well on top of a sodium dodecyl sulfate (SDS)-polyacrylamide gel. This second gel separates the proteins according to their mass. In this diagram, only a few spots are seen, but an actual experiment would involve a mixture of hundreds or thousands of different proteins. (**b**) A photograph of an SDS-polyacrylamide gel that has been stained for proteins. Each spot represents a unique cellular protein.

is termed **isoelectric focusing.** After the tube gel has run, it is laid horizontally on top of a polyacrylamide slab gel that contains sodium dodecyl sulfate (SDS). The SDS coats each protein with negative charges and denatures them. Proteins in the slab gel are separated according to their molecular mass. Smaller proteins move toward the bottom of the gel more quickly than larger ones. After the slab gel has run, the proteins within the gel can be stained with a dye. As seen in Figure 22.4b, the end-result is a collection of spots, each spot corresponding to a unique cellular protein, with proteins of a larger mass remaining higher in the gel. The resolving power of two-dimensional gel electrophoresis is extraordinary. Proteins that differ by a single charged amino acid can be resolved as two distinct spots using this method.

Various approaches can be followed to identify spots on a gel that may be of interest to researchers. One possibility is that a given cell type may show a few very large spots that are not found when proteins are analyzed from other cell types. The relative abundance of such spots may indicate that a particular protein is important to that cell's structure or function. Secondly, certain spots on a two-dimensional gel may be seen only under a given set of conditions. For example, a researcher may be interested in the effects of a hormone on the function of a particular cell type. Two-dimensional gel electrophoresis could be conducted on a sample in which the cells had not been exposed to the hormone versus a sample in which they had. This may reveal particular spots that are present only when the cells are exposed to a given hormone. Finally, abnormal cells often express proteins that are not found in normal cells. This is particularly true for cancer cells. A researcher may compare normal and cancer cells via two-dimensional gel electrophoresis to identify proteins expressed only in cancer cells.

Mass Spectrometry Is Used to Identify Proteins

Two-dimensional gel electrophoresis may be used as the first step in the separation of cellular proteins. The next goal is to correlate a given spot on a two-dimensional gel with a particular protein. To accomplish this goal, a spot on a two-dimensional gel can be cut out of the gel to obtain a tiny amount of the protein within the spot. In essence, the two-dimensional gel electrophoresis procedure purifies a small amount of the cellular protein of interest. The next step is to determine the amino acid sequence of a region of the protein, which aids in its identification. This can be accomplished via **mass spectrometry,** a technique that measures the mass of a molecule, such as a peptide fragment.

As shown in Figure 22.5, the technique actually determines the mass of peptide fragments that are produced by digesting a purified protein with a protease that cuts the protein into small peptide fragments. The peptides are mixed with an organic acid and dried onto a metal slide. The sample is then struck with a laser. This ejects the peptides from the slide in the form of an ionized gas in which the peptide contains one or more positive charges. The charged peptides are then accelerated via an electric field and fly toward a detector. The time of flight is determined by their mass and net charge. A measurement of the flight time provides an extremely accurate way of determining the mass of a peptide.

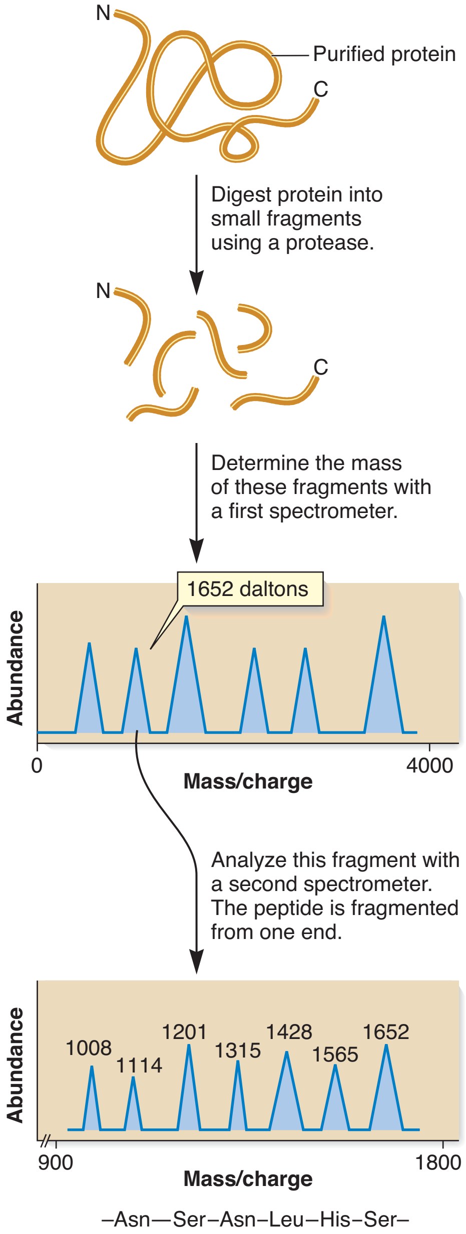

N
Purified protein
C
Digest protein into small fragments using a protease.
N
C
Determine the mass of these fragments with a first spectrometer.
1652 daltons
Abundance
0
4000
Mass/charge
Analyze this fragment with a second spectrometer. The peptide is fragmented from one end.
Abundance
1008
1114
1201
1315
1428
1565
1652
900
1800
Mass/charge
–Asn—Ser–Asn–Leu–His–Ser–

FIGURE 22.5 **The use of tandem mass spectrometry to determine the amino acid sequence of a peptide.**

Concept check: What is the purpose of tandem mass spectrometry?

An ultimate goal of mass spectrometry is to determine the amino acid sequence of a given peptide. This is accomplished using two mass spectrometers, a method called **tandem mass spectrometry.** The first mass spectrometer measures the mass of a given peptide. This same peptide is then analyzed by a second spectrometer after the peptide has been broken down into smaller fragments. The differences in the masses of the peaks in the spectrum from the second spectrometer reveal the amino acid sequence of the peptide, because the masses of all 20 amino acids are known. For example, as shown in Figure 22.5, let's suppose a peptide had a mass of 1652 daltons (Da). If one amino acid at the end was removed and the smaller peptide had a mass that was 87 Da less (i.e., 1565 Da), this indicates that a serine is at one end of the peptide, because the mass of serine within a polypeptide chain is 87 Da. Removing two amino acids at one end produces a mass that is 224 Da less, corresponding to the removal of one serine (87 Da) and one histidine (137 Da). If three amino acids were removed and the mass was decreased by 337 Da, this corresponds to the removal of one serine (87 Da), one histidine (137 Da), and one leucine (113 Da). Therefore, from these measurements, we conclude that the amino acid sequence from one end of the peptide was serine-histidine-leucine.

How does this information lead to the identification of a specific protein? After a researcher has obtained a few short peptide sequences from a given protein, researchers can compare their results with information that has already been obtained and stored in large databases. Using computer software described later in this chapter, the amino acid sequences obtained by mass spectrometry can be used as query sequences to search a large database that contains a collection of protein sequences. This program may locate a match between the experimental sequences and a specific protein within a particular species. In this way, mass spectrometry makes it possible to identify the purified protein.

Mass spectrometry can also identify protein covalent modifications. For example, if an amino acid within a peptide was phosphorylated, the mass of the peptide is increased by the mass of a phosphate group. This increase in mass can be determined via mass spectrometry.

Protein Microarrays Can Be Used to Study Protein Expression and Function

Earlier in this chapter, we learned about DNA microarrays, which have gained widespread use to study gene expression at the RNA level. The technology for making DNA microarrays is also being applied to make **protein microarrays.** In this type of technology, proteins, rather than DNA molecules, are spotted onto a glass or silica slide. The development of protein microarrays is more challenging because proteins are much more easily damaged by the manipulations that occur during microarray formation. For example, the three-dimensional structure of a protein may be severely damaged by drying, which usually occurs during the formation of a microarray. This has created additional challenges for researchers who are developing the technology of protein microarrays. In addition, the synthesis and purification of proteins tend to be more time-consuming than the production of DNA, which can be amplified by PCR or directly synthesized on the microarray itself. In spite of these technical difficulties, the last few years have seen progress in the production and uses of protein microarrays (Table 22.2).

The two common types of protein microarray analyses are antibody microarrays and functional protein microarrays. The purpose of an **antibody microarray** is to study protein expression. Antibodies are proteins that recognize antigens. One type of antigen that an antibody can recognize is a short peptide sequence found within another protein. Therefore, an antibody can specifically recognize a cellular protein. Researchers and commercial laboratories can produce different antibodies; each antibody recognizes a specific peptide sequence. These can be spotted onto a microarray. Cellular proteins can then be isolated, fluorescently labeled, and exposed to the antibody microarray. When a given protein is recognized by an antibody on the microarray, it is captured by the antibody and remains bound to that spot. The level of fluorescence at a given spot indicates the amount of a cellular protein that is recognized by a particular antibody.

TABLE **22.2**

Some Applications of Protein Microarrays

Application	Description
Protein expression	An antibody microarray can measure protein expression because each antibody in a given spot recognizes a specific amino acid sequence. This can be used to study the expression of proteins in a cell-specific manner. It can also be used to determine how environmental conditions affect the levels of particular proteins.
Protein function	The substrate specificity and enzymatic activities of groups of proteins can be analyzed by exposing a functional protein microarray to a variety of substrates.
	Protein-protein interactions The ability of two proteins to interact with each other can be determined by exposing a functional protein microarray to fluorescently labeled proteins.
Pharmacology	The ability of drugs to bind to cellular proteins can be determined by exposing a functional protein microarray to different kinds of labeled drugs. This can help to identify the proteins within a cell to which a given drug may bind.

The other type of array is a **functional protein microarray.** To make this type of array, researchers must purify cellular proteins and then spot them onto a microarray. The microarray can then be analyzed with regard to specific kinds of protein function. In 2000, for example, Heng Zhu, Michael Snyder, and colleagues purified 119 proteins from yeast that were known to function as protein kinases. These kinds of proteins attach phosphate groups onto other cellular proteins. A microarray was made consisting of different possible proteins that may or may not be phosphorylated by these 119 kinases, and then the array was exposed to each of the kinases in the presence of radiolabeled ATP. By following the incorporation of phosphate into the array, they determined the protein specificity of each kinase. On a much larger scale, the same group of researchers purified 5800 different yeast proteins and spotted them onto a microarray. The array was then exposed to fluorescently labeled calmodulin, which is a regulatory protein that binds calcium ions. Several proteins in the microarray were found to bind calmodulin. Although some of these were already known to be regulated by calmodulin, other proteins in the array that had not been previously known to bind calmodulin were identified.

22.2 REVIEWING THE KEY CONCEPTS

- The proteome is the entire collection of proteins that a cell or organism makes. The study of the function and interactions of many proteins is called proteomics.
- The proteome is much larger than the genome due to alternative splicing, RNA editing, and posttranslational covalent modifications (see Figure 22.3).
- Two-dimensional gel electrophoresis is used to separate a complex mixture of proteins (see Figure 22.4).
- Mass spectrometry and tandem mass spectrometry are used to identify short amino acids sequences within purified proteins. These short sequences can be used to identify a protein (see Figure 22.5).
- Protein microarrays are used to study protein expression, protein function, protein-protein interactions, and protein-drug interactions (see Table 22.2).

22.2 COMPREHENSION QUESTIONS

1. Which of the following is a reason why the proteome of a eukaryotic cell is usually much larger than its genome?
   a. Alternative splicing
   b. RNA editing
   c. Posttranslational covalent modifications
   d. All of the above
2. During two-dimensional gel electrophoresis, proteins are separated based on
   a. their net charge at a given pH.
   b. their mass.
   c. their ability to bind to a specific resin.
   d. both a and b.
3. The technique of tandem mass spectrometry is used to determine
   a. the amino acid sequence of a peptide fragment.
   b. the nucleotide sequence of a segment of RNA.
   c. the nucleotide sequence of a segment of DNA.
   d. the number of genes in a species' genome.
4. Which of the following can be analyzed using a protein microarray?
   a. Protein expression
   b. Protein function
   c. Protein-protein interactions
   d. All of the above

22.3 BIOINFORMATICS

Geneticists use computers to collect, store, manipulate, and analyze data. Molecular genetic data is particularly amenable to computer analysis because it comes in the form of a sequence, such as a DNA, RNA, or amino acid sequence. The ability of computers to analyze data at a rate of millions or even billions of operations per second has made it possible to solve problems concerning genetic information that were thought intractable a few decades ago.

In recent years, the marriage between genetics and computational tools and approaches has yielded an important branch of science known as bioinformatics. Several scientific journals are largely devoted to this topic. Computer analysis of genetic sequences relies on three basic components: a computer, a computer program, and some type of data. In genetic research, the data consist of a particular genetic sequence or several sequences that a researcher or clinician wants to study. For example, this could be a DNA sequence derived from a cloned DNA fragment. A sequence of interest may be relatively short or thousands to millions of nucleotides in length. Experimentally, DNA sequences and related data are obtained using the techniques described in Chapters 19 and 21.

In this section, we first consider the fundamental concepts that underlie the analysis of genetic sequences. We then explore how these methods are used to provide insights regarding functional genomics and proteomics. Chapter 27 will describe applications of bioinformatics in the area of evolutionary genetics. In addition, you may wish to actually run computer programs, which are widely available at university and government websites (e.g., see www.ncbi.nlm.nih.gov/Tools). This type of hands-on learning will help you see how the computer has become a valuable tool for analyzing genetic data.

Sequence Files Are Analyzed by Computer Programs

Most people are familiar with **computer programs,** which consist of a defined series of operations that can manipulate and analyze information in a desired way. For example, a computer program might be designed to take a DNA sequence and translate it into an amino acid sequence. A first step in the computer analysis of genetic data is the creation of a **computer data file** to store the data. This file is simply a collection of information in a form suitable for storage and manipulation on a computer. In genetic studies, a computer data file might contain an experimentally obtained DNA, RNA, or amino acid sequence. For example, a file could contain the DNA sequence of one strand of the *lacY* gene from *Escherichia coli*, as shown here. The numbers to the left represent the base number in the sequence file.

```
    1 ATGTACTATT TAAAAAACAC AAACTTTTGG ATGTTCGGTT TATTCTTTTT
   51 CTTTTACTTT TTTATCATGG GAGCCTACTT CCCGTTTTTC CCGATTTGGC
  101 TACATGACAT CAACCATATC AGCAAAGTG  ATACGGGTAT TATTTTGCC
  151 GCTATTTCTC TGTTCTCGCT ATTATTCCAA CCGCTGTTTG GTCTGCTTC
  201 TGACAAACTC GGGCTGCGCA AATACCTGCT GTGGATTATT ACGGCATGT
  251 TAGTCATGTT TGCGCCGTTC TTTATTTTTA TCTTCGGGCC ACTGTTACAA
  301 TACAACATTT TAGTAGGATC GATTGTTGGT GGTATTTATC TAGGCTTTTG
  351 TTTTAACGCC GGTGCGCCAG CAGTAGAGGC ATTTATTGAG AAAGTCAGCC
  401 GTCGCAGTAA TTTCGAATTT GGTCGCGCGC GGATGTTTGG CTGTGTTGGC
  451 TGGGCGCTGT GTGCCTGAT  TGTCGGCATC ATGTTCACCA TCAATAATCA
  501 GTTTGTTTTC TGGCTGGGCT CTGGCTGTGC ACTCATCCTC GCCGTTTTAC
  551 TCTTTTTCGC CAAAACGGAT GCGCCCTCTT CTGCCACGGT TGCCAATGCG
  601 GTAGGTGCCA ACCATTCGGC ATTTAGCCTT AAGTGGCAC  TGGAACTGTT
  651 CAGACAGCCA AAACTGTGGT TTTGTCTACT GTATGTTATT GGCGTTTCCT
  701 GCACCTACGA TGTTTTTGAC CAACAGTTTG CTAATTTCTT TACTTCGTTC
  751 TTTGCTACCG GTGAACAGGG TACGCGGGTA TTTGGCTACG TAACGACAAT
  801 GGGCGCAATTA CTTAACGCCT CGATTATCTT CTTGCGCCA CTGATCATTA
  851 ATCGCATCGG TGGGAAAAAC GCCCTGCTGC TGGCTGGCAC TATTATCTCT
  901 CTACGTATTA TTGGCTCATC GTTCGCCACC TCAGCGCTGG AAGTGGTTAT
  951 TCTGAAAACG CTGCATATGT TTGAAGTACC GTTCCTGCTG GTGGGCTGCT
1,001 TTAAATATAT TACCAGCCAG TTTGAAGTGC GTTTTTCAGC GACGATTTAT
1,051 CTGGTCTGTT TCTGCTTCTT TAAGCAACTG GCATGATTT  TTATGTCTGT
1,101 ACTGGCGGGC AATATGTATG AAAGCATCGG TTTCCAGGGC GCTTATCTGG
1,151 TGCTGGGTCT GGTGGCGCTG GGCTTCACCT TAATTTCCGT GTTCACGCTT
1,201 AGCGGCCCCG GCCCGCTTTC CCTGCTGCGT CGTCAGGTGA ATGAAGTCGC
1,251 TTAA
```

To store data in a computer data file, a scientist creates the file and enters the data, either by hand or, as is now more common, by laboratory instruments such as densitometers and fluorometers. These instruments have the capability to read data, such as a sequencing ladder, and enter the DNA sequence information directly into a computer file.

The purpose of making a computer file that contains a genetic sequence is to take advantage of the swift speed with which computers can analyze this information. Genetic sequence data in a computer file can be investigated in many different ways, corresponding to the many questions a researcher might ask about the sequence and its functional significance. The possibilities include the following:

1. Does a sequence contain a gene?
2. Where are functional sequences such as promoters, regulatory sites, and splice sites located within a particular gene?
3. Does a sequence encode a polypeptide? If so, what is the amino acid sequence of the polypeptide?
4. Does a sequence predict certain structural features for DNA, RNA, or proteins? For example, is a DNA sequence likely to be in a Z-DNA conformation? What is the secondary structure of an RNA sequence or a polypeptide sequence?
5. Is a sequence homologous to any other known sequences?
6. What is the evolutionary relationship between two or more genetic sequences?

To answer these and many other questions, different computer programs have been written to analyze genetic sequences in particular ways. These programs have been devised by theoreticians who understand basic genetic principles and can design computational strategies for analyzing genetic sequences. When constructing a computer program, a theoretician has a goal that the program is meant to fulfill. For example, a theoretician may wish to write a program to translate a DNA sequence into an amino acid sequence. Based on knowledge of the genetic code, she or he can devise computational procedures, or algorithms, that relate a DNA sequence to an amino acid sequence. In a computer program, these procedures are executed as a stepwise "plan of operations" that manipulates the data in a sequence file.

As an example, let's consider a computer program aimed at translating a DNA sequence into an amino acid sequence and consider how it might work in practice. The geneticist—the user—wants to translate a DNA sequence file into an amino acid sequence. The user is sitting at a computer that is connected to a program that can translate a DNA sequence into an amino acid sequence. The program is opened and the user provides the program with a DNA sequence file to be translated. (See the sequence of the *lacY* gene on the previous page.) The user also specifies which portion of the sequence should be translated. In this case, the user decides to begin the translation at the first nucleotide in the sequence file and end the translation at nucleotide number 1254. The user wants the program to translate the sequence in all three forward reading frames and to show the longest reading frame—the longest amino acid sequence that is uninterrupted by a stop codon. This translated sequence is saved in a file. The contents of the file are shown here.

```
  1 MYYLKNTNFW MFGLFFFFYF FIMGAYFPFF PIWLHDINHI SKSDTGIIFA
 51 AISLFSLLFQ PLFGLLSDKL GLRKYLLWII TGMLVMFAPF FIFIFGPLLQ
101 YNILVGSIVG GIYLGFCFNA GAPAVEAFIE KVSRRSNFEF GRARMFGCVG
151 WALCASIVGI MFTINNQFVF WLGSGCALIL AVLLFFAKTD APSSATVANA
201 VGANHSAFSL KLALELFRQP KLWFLSLYVI GVSCTYDVFD QQFANFFTSF
251 FATGEQGTRV FGYVTTMGEL LNASIMFFAP LIINRIGGKN ALLLAGTIMS
301 VRIIGSSFAT SALEVVIKLT LHMFEVPFLL VGCFKYITSQ FEVRFSATIY
351 LVCFCFFKQL AMIFMSVLAG NMYESIGFQG AYLVLGLVAL GFTLISVFTL
401 SGPGPLSLLR RQVNEVA
```

In this file, which was created by a computer program, each of the 20 amino acids is given a single-letter abbreviation (see Figure 15.5).

Why is such a program useful? The advantages of running this program are speed and accuracy. It can translate a relatively long genetic sequence within milliseconds. By comparison, it would probably take you a few hours to look each codon up in the genetic code table and write the sequence out in the correct order. If you visit a website and run a program like this, you will discover that such a program can translate a genetic sequence into six reading frames—three forward and three reverse. This is useful if a researcher does not know where the start codon is located and/or does not know the direction of the coding sequence.

In genetic research, large software packages typically contain many computer programs that can analyze genetic sequences in different ways. For example, one program can translate a DNA sequence into an amino acid sequence, and another program can locate introns within genes. These software packages are found at universities, government facilities, hospitals, and industries. At such locations, a central computer with substantial memory and high-speed computational abilities runs the software, and individuals can connect to this central computer via their personal computer (e.g., laptop). Many such programs are freely available on the Internet. See www.ncbi.nlm.nih.gov/Tools for a variety of useful programs.

The Scientific Community Has Collected Sequence Files and Stored Them in Large Computer Databases

In Chapter 19, we considered how researchers can clone and sequence genes, and in Chapter 21, we learned how scientists are investigating the genetic sequences of entire genomes from several species, including humans. The amount of genetic information generated by researchers has become enormous. For example, at its completion, the Human Genome Project had produced more data than any other undertaking in the history of biology. With these advances, scientists realize that another critical use of computers is simply to store the staggering amount of data produced from genetic research.

A large number of computer data files collected and stored in a single location is called a **database.** In addition to genetic sequences, the files within databases are **annotated,** which means they contain additional information such as a concise description of a sequence, the name of the organism from which this sequence was obtained, and the function of the encoded protein, if it is known. The file may also describe other features of significance and cite a published journal reference that describes the sequence.

The scientific community has collected the genetic information from thousands of research labs and created several large databases. Table 22.3 describes some of the major genetic databases in use worldwide. These databases enable researchers to access and compare genetic sequences that are obtained by many laboratories.

TABLE 22.3

Examples of Major Computer Databases

Type	Description
Nucleotide sequence	DNA sequence data are collected into three internationally collaborating databases: GenBank (a U.S. database), EMBL (European Molecular Biology Laboratory Nucleotide Sequence Database), and DDBJ (DNA Databank of Japan). These databases receive sequence and sequence annotation data from genome projects, sequencing centers, individual scientists, and patent offices. These databases are accessed via the Internet.
Amino acid sequence	Amino acid sequence data are collected into a few international databases, including Swissprot (Swiss protein database), PIR (Protein Information Resource), Genpept (translated peptide sequences from the GenBank database), and TrEMBL (Translated sequences from the EMBL database).
Three-dimensional structure	PDB (Protein Data Bank) collects the three-dimensional structures of biological macromolecules with an emphasis on protein structure. These are primarily structures that have been determined by X-ray crystallography and nuclear magnetic resonance (NMR), but some models are included in the database. These structures are stored in files that can be viewed on a computer with the appropriate software.
Protein motifs	Prosite is a database containing a collection of amino acid sequence motifs that are characteristic of a protein family, domain structure, or certain posttranslational modifications. Pfam is a database of protein families with multiple amino acid sequence alignments.
Gene expression data	Gene Expression Omnibus (GEO) contains data regarding the expression patterns of genes within a data set, such as the data obtained from microarrays or ChIP-chip assays.

The databases described in Table 22.3 collect genetic information from many different species. Scientists have also made more specialized databases, called **genome databases,** that focus on the genetic sequences and characteristics of a single species. Genome databases have been created for species of bacteria (e.g., *E. coli*), yeast (e.g., *Saccharomyces cerevisiae*), worms (e.g., *Caenorhabditis elegans*), fruit flies (e.g., *Drosophila melanogaster*), plants (e.g., *Arabidopsis thaliana*), and mammals (e.g., mice and humans). The primary aim of genome databases is to organize the information from sequencing and mapping projects for a single species. Genome databases identify an organism's known genes and describe their map location within the genome. In addition, a genome database may provide information concerning gene alleles, bibliographic information, and other pertinent information.

Different Computational Strategies Can Identify Functional Genetic Sequences

At the molecular level, the function of the genetic material is based largely on specific genetic sequences that play distinct roles. For example, codons are three-base sequences that specify particular amino acids, and promoters are sequences that provide a binding site for RNA polymerase to initiate transcription. Computer programs can be designed to scan very long sequences, such as those obtained from genome-sequencing projects, and locate features of interest within them. To illustrate this concept, let's first consider the following sequence file, which contains an alphabetic sequence of 54 letters:

Sequence file:

GJTRLLAMAQLHEOGYLTOBWENTMNMTORXXXTGOODNTH
EQALLRTLSTORE

Now let's compare how three different computer programs might analyze this sequence to identify meaningful features. The goal of our first program is to locate all of the English words within this sequence. If we ran this program, we would obtain the following result:

GJTRLLAMAQLHEOGYLTOBWENTMNMTORXXXTGOODN
THEQALLRTLSTORE

In this case, a computer program has identified locations where the sequence of letters forms a word. Several words (which are underlined) have been located within this sequence.

A second computer program could be aimed at locating a series of words that are organized in the correct order to form a grammatically logical English sentence. If we used our sequence file and ran this program, we would obtain the following result:

GJTRLLAMAQLHEOGYLTOBWENTMNMTORXXXTGOODN
THEQALLRTLSTORE

The second program has identified five words that form a logical sentence.

Finally, a computer program might be used to identify patterns of letters, rather than words or sentences. For example, a computer program could locate a pattern of five letters that occurs in both the forward and reverse directions. If we applied this program to our sequence file, we would obtain the following:

GJTRLLAMAQLHEOGYLTOBWENTMNMTORXXXTGOODNTH
EQALLRTLSTORE

In this case, the program has identified a pattern where five letters are found in both the forward and reverse directions.

In the three previous examples, we can distinguish between **sequence recognition** (as in our first example) and **pattern recognition** (as in our third example). In sequence recognition, the computer program has the information that a specific sequence of symbols has a specialized meaning. This information must be supplied to the computer program. For example, the first program has access to the information from a dictionary with all known English words. With this information, the first program can identify sequences of letters that make words. By comparison, the third computer program does not rely on specialized sequence information. Rather, it is looking for a pattern of symbols that can occur within any group of symbol arrangements.

Overall, the simple programs we have considered illustrate three general types of identification strategies:

1. *Locate specialized sequences within a very long sequence.* A specialized sequence with a particular meaning or function is called a **sequence element,** or **motif.** The computer program has a list of predefined sequence elements and can identify such elements within a sequence of interest.
2. *Locate an organization of sequences.* As shown in the second computer program, this could be an organization of sequence elements. Alternatively, it could be an organization of a pattern of symbols.
3. *Locate a pattern of symbols.* The third computer program is an example of locating a pattern of symbols.

The great power of computer analysis is that these types of operations can be performed with great speed and accuracy on sequences that may be enormously long.

Now that we understand the general ways that computer programs identify sequences, let's consider specific examples. As we have discussed throughout this textbook, many short nucleotide sequences play specialized roles in the structure or function of genetic material. A geneticist may want to locate a short sequence element within a longer nucleotide sequence in a data file. For example, a sequence of chromosomal DNA might be tens of thousands of nucleotides in length, and a geneticist may want to know whether a sequence element, such as a TATA box, is found at one or more sites within the chromosomal DNA. To do so, a researcher could visually examine the long chromosomal DNA sequence in search of a TATA box. Of course, this process would be tedious and prone to error. By comparison, the appropriate computer program can locate a sequence element in seconds. Therefore, computers are very useful for this type of application. **Table 22.4** lists some examples of sequence elements that can be identified by computer analysis.

By comparing the amino acid sequences and known functions of proteins in thousands of cases, researchers have also found protein motifs that carry out specialized functions within proteins. For example, researchers have determined that the amino acid sequence motif asparagine-X-serine (where X is any amino acid except proline) within eukaryotic proteins is a glycosylation site (i.e., it may have a carbohydrate attached to it). The Prosite database contains a collection of all amino acid sequence motifs known to be functionally important. Researchers can use computer programs to determine whether an amino acid sequence contains any of the motifs found in the Prosite database. This may help them to understand the role of a newly found protein of unknown function.

TABLE **22.4**

Short Sequence Elements That Can Be Identified by Computer Analysis

Type of Sequence	Examples*
Promoter	Many *E. coli* promoters contain TTGACA (−35 site) and TATAAT (−10 site). Eukaryotic core promoters may contain TATA boxes, etc.
Response elements	Glucocorticoid response element (AGRACA)
Start codon	ATG
Stop codons	TAA, TAG, TGA
Splice site	GTRAGT————YNYTRAC(Y)$_n$AG
Polyadenylation signal	AATAAA
Highly repetitive sequences	Relatively short sequences that are repeated many times throughout a genome
Transposable elements	Often characterized by a pattern in which direct repeats flank inverted repeats

*The sequences shown in this table would be found in the DNA. For gene sequences, only the coding strand is shown. R = purine (A or G); Y = pyrimidine (T or C); N = A, T, G, or C; U in RNA = T in DNA.

Homologous Genes Are Derived from the Same Ancestral Gene

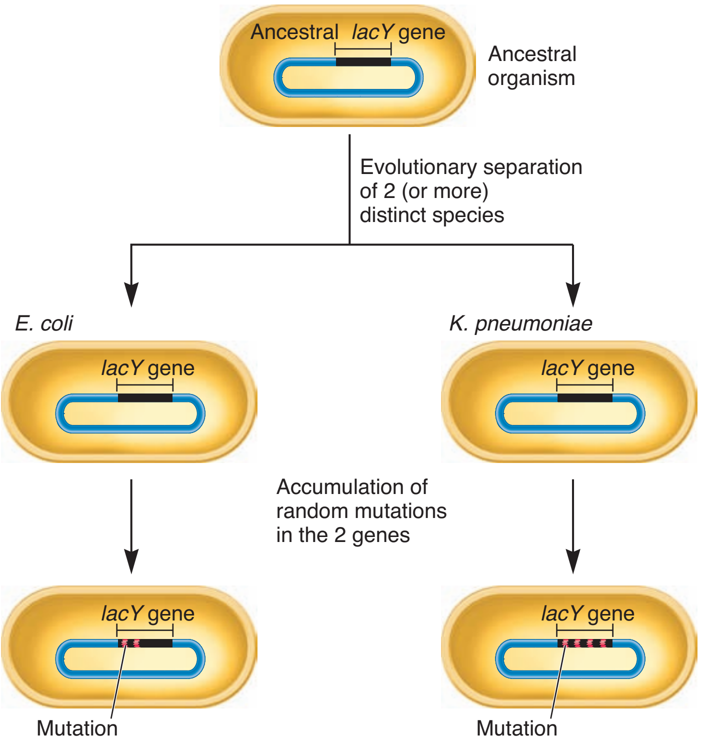
Computer technology is also used to identify genes that are evolutionarily related. This has become an extremely powerful tool in the field of bioinformatics. Before we discuss this computer technology, let's first consider how the sequences of different genes may be related to each other. When comparing gene sequences, researchers frequently find two or more similar sequences. For example, the sequence of the *lacY* gene that encodes the lactose permease in *E. coli* is similar to that of the *lacY* gene that encodes the lactose permease in another bacterium, *Klebsiella pneumoniae*. As shown in **Figure 22.6a**, when segments of the two *lacY* genes are lined up, approximately 78% of their bases are a perfect match.

In this case, the two sequences are similar because the genes are **homologous,** meaning they have been derived from the same ancestral gene. This idea is shown schematically in **Figure 22.6b**. An ancestral *lacY* gene was located in a bacterium that preceded the evolutionary divergence between *E. coli* and *K. pneumoniae*. After these two species diverged from each other, their *lacY* genes accumulated distinct mutations that produced somewhat different base sequences for this gene. Therefore, in these two species of bacteria, the *lacY* genes are similar but not identical. When two homologous genes are found in different species, they are termed **orthologs.**

```
              151                                                        200
E. coli       TCTTTTTCTT  TTACTTTTTT  ATCATGGGAG  CCTACTTCCC  GTTTTTCCCG
K. pneumoniae TCTTTTTCTT  TTACTATTTC  ATTATGTCAG  CCTACTTTCC  TTTTTTTCCG

              201                                                        250
E. coli       ATTTGGCTAC  ATGACATCAA  CCATATCAGC  AAAAGTGATA  CGGGTATTAT
K. pneumoniae GTGTGGCTGG  CGGAAGTTAA  CCATTTAACC  AAAACCGAGA  CGGGTATTAT
```

(a) A comparison of a portion of the *lacY* gene from *E. coli* and *K. pneumoniae*



(b) The formation of homologous genes, also known as orthologs

FIGURE 22.6 The origin of homologous *lacY* genes in *Escherichia coli* and *Klebsiella pneumoniae*. (**a**) A comparison of a short region of the *lacY* gene from *E. coli* and *K. pneumoniae*. (**b**) This figure emphasizes a single gene within an ancestral organism. During evolution, the ancestral organism diverged into two different species: *E. coli* and *K. pneumoniae*. After this divergence, the *lacY* gene in the two separate species accumulated mutations, yielding *lacY* genes with somewhat different sequences.

Genes→Traits After two species diverge evolutionarily, their genes accumulate different random mutations. This example concerns the *lacY* gene, which encodes lactose permease. In both species, the function of lactose permease is to transport lactose into the cell. The *lacY* gene in these two species has accumulated different mutations that alter the amino acid sequence of the protein. Researchers have determined that these two species transport lactose at significantly different rates. Therefore, the changes in gene sequences have affected the ability of these two species to transport lactose.

Concept check: Explain why the sequences of these two genes are similar to each other but not identical.

Two or more homologous genes can also be found within a single species. These are termed paralogous genes, or **paralogs.** As discussed in Chapter 8, abnormal gene duplications may happen several times during evolution, which results in multiple copies of a gene and ultimately leads to the formation of a gene family. A **gene family** consists of two or more paralogs within the genome of a single species. When a gene family occurs, the concept of orthologs becomes more complex. For example, let's consider the globin gene family found in mammals (see Chapter 8, Figure 8.7). Researchers would say that the β-globin gene in humans is an ortholog to the β-globin gene found in mice. Likewise, α-globin genes found in both species are considered orthologs. However, the α-globin gene in humans is not called an ortholog of the β-globin gene in mice, although they could be called homologous. The most closely related genes in two different species are considered orthologs.

Homologous genes, whether they are orthologs or paralogs, have similar sequences. What is the distinction between homology and sequence similarity? **Homology** implies a common ancestry. **Similarity** means that two sequences are similar to each other. In many cases, such as the *lacY* example, similarity is due to homology. However, this is not always the case. Short genetic sequences may be similar to each other even though two genes are not related evolutionarily. For example, many nonhomologous bacterial genes contain similar promoter sequences at the −35 and −10 regions.

A Database Can Be Searched to Identify Homologous Sequences

Homologous genes usually encode proteins that carry out similar or identical functions. As we have just considered, the *lacY* genes in *E. coli* and *K. pneumoniae* both encode lactose permease, a protein that transports lactose across the bacterial cell membrane.

A strong correlation is typically found between homology and function. How is this relationship useful with regard to bioinformatics? In many cases, the first indication of the function of a newly determined sequence is through homology to known sequences in a database. An example is the *CFTR* gene that is altered in cystic fibrosis patients. After the *CFTR* gene was identified in humans, a database search revealed it is homologous to several genes found in other species. Moreover, a few of the homologous genes were already known to encode proteins that function in the transport of ions and small molecules across the plasma membrane. This observation provided an important clue that cystic fibrosis involves a defect in ion transport.

The ability of computer programs to identify homology between genetic sequences provides a powerful tool for predicting the function of genetic sequences. In 1990, Stephen Altschul, David Lipman, and colleagues developed a program called **BLAST** (for **basic local alignment search tool**). The BLAST program has been described by many geneticists as the single most important bioinformatic tool available. This computer program starts with a particular genetic sequence and then locates homologous sequences within a large database.

To see how the BLAST program works, let's consider the human enzyme phenylalanine hydroxylase, which functions in

TABLE 22.5

Results from a BLAST Program Comparing Human Phenylalanine Hydroxylase with Database Sequences

Match	% of Identical Amino Acids*	Species	Function of Sequence†	E-value
1	100	Human (*Homo sapiens*)	Phenylalanine hydroxylase	0
2	99	Orangutan (*Pongo pygmaeus*)	Phenylalanine hydroxylase	0
3	92	Mouse (*Mus musculus*)	Phenylalanine hydroxylase	0
4	92	Rat (*Rattus norvegicus*)	Phenylalanine hydroxylase	0
5	83	Chicken (*Gallus gallus*)	Phenylalanine hydroxylase	0
6	78	Western clawed frog (*Xenopus tropicalis*)	Phenylalanine hydroxylase	0
7	75	Zebrafish (*Danio rerio*)	Phenylalanine hydroxylase	0
8	72	Japanese pufferfish (*Takifugu rubripes*)	Phenylalanine hydroxylase	0
9	62	Fruit fly (*Drosophila melanogaster*)	Phenylalanine hydroxylase	10^{-154}
10	57	Nematode (*Caenorhabditis elegans*)	Phenylalanine hydroxylase	10^{-141}

*The number indicates the percentage of amino acids that are identical to the amino acid sequence of human phenylalanine hydroxylase.

†In some cases, the function of the sequence was determined by biochemical assay. In other cases, the function was inferred due to the high degree of sequence similarity with other species.

the metabolism of phenylalanine, an amino acid. Recessive mutations in the gene that encodes this enzyme are responsible for the disease called phenylketonuria (PKU). The computational experiment shown in **Table 22.5** started with the amino acid sequence of this protein in humans and used the BLAST program to search the Swissprot database, which contains millions of different protein sequences. The BLAST program can determine which sequences in this database are the closest matches to the amino acid sequence of human phenylalanine hydroxylase. Table 22.5 shows a portion of the results—10 selected matches to human phenylalanine hydroxylase that were identified by the program. Because this enzyme is found in nearly all eukaryotic species, the program identified phenylalanine hydroxylase from many different species. The column to the right of the match number shows the percentage of amino acids that are identical between the species indicated and the human sequence. Because the human phenylalanine hydroxylase sequence is already in the Swissprot database, the closest match of human phenylalanine hydroxylase is to itself (100%). The next nine sequences are in order of similarity. The next most similar sequence is from the orangutan (99%), a close relative of humans. This is followed by two mammals, the mouse and rat, and then four vertebrates that are not mammals. The ninth and tenth best matches are from *Drosophila* and *C. elegans*, which are invertebrates.

As shown in the right column of Table 22.5, the relationship between the query sequence and each matching sequence is given an E-value (Expect value). The **E-value** represents the number of times that the match or a better one would be expected to occur purely by random chance in a search of the entire database. An E-value that is very small indicates that the similarity between the query sequence and the matching sequence is unlikely to have occurred by random chance. Instead, researchers would accept the hypothesis that the two sequences are homologous, which means they are derived from the same ancestral sequence.

E-values are very dependent on different parameters such as the length of the query sequence, the number of gaps that are needed to align the query sequence and the matching sequence, and the database size. As a general rule, if the E-value is less than 1×10^{-50}, the match is very similar to the query sequence and is likely to be homologous. (Values that are much less than 10^{-100} are reported as zero by the BLAST program.) If the value lies between 1×10^{-50} and 1×10^{-10}, the match, or part of it, is likely to be homologous. If the value is between 1×10^{-10} and 1×10^{-2}, the match has a significant chance of being related to the query sequence, whereas values between 1×10^{-2} and 1 have a relatively low probability of being homologous. Values above 1 are usually not evolutionarily related. As seen in Table 22.5, all of the E-values are below 1×10^{-140}, which suggests that all of these matches are homologous to the query sequence.

You can see two trends in Table 22.5. First, the order of the matches follows the evolutionary relatedness of the various species to humans. The similarity between any two sequences is related to the time that has passed since they diverged from a common ancestor. Among the species listed in this table, the human sequence is most similar to the orangutan, a closely related primate. The next most similar sequences are found in other mammals, followed by other vertebrates, and finally invertebrates. A second trend you may have noticed is that several of the matches involve species that are important from a research, medical, or agricultural perspective. Currently, our genetic databases are biased toward organisms that are of interest to humans, particularly model organisms such as mice and *Drosophila*. Over the next several decades, the sequencing of genomes from many different species will tend to lessen this bias.

The results shown in Table 22.5 illustrate the remarkable computational abilities of current computer technology. In minutes, the amino acid sequence of human phenylalanine hydroxylase can be compared with millions of different sequences.

Genetic Sequences Can Be Used to Predict the Structure of RNA and Proteins

Another topic in which bioinformatics has influenced functional genomics and proteomics is the area of structure prediction. The function of macromolecules such as DNA, RNA, and proteins relies on their three-dimensional structure, which, in turn, depends on the linear sequences of their building blocks. In the case of DNA and RNA, this means a linear sequence of nucleotides; proteins are composed of a linear sequence of amino acids. Currently, the three-dimensional structure of macromolecules is determined primarily through the use of biophysical techniques such as X-ray crystallography and nuclear magnetic resonance (NMR). These methods are technically difficult and very time-consuming. DNA sequencing, by comparison, requires much less effort. Therefore, because the three-dimensional structure of macromolecules depends ultimately on the linear sequence of their building blocks, it would be far easier if we could predict the structure (and function) of DNAs, RNAs, and proteins from their sequence of building blocks.

RNA molecules typically are folded into a secondary structure, which commonly contains double-stranded regions. This secondary structure is further folded and twisted to adopt a tertiary conformation. Such structural features of RNA molecules are functionally important. For example, the folding of RNA into secondary structures, such as stem-loops, affects transcriptional termination and other regulatory events. Therefore, geneticists are interested in the secondary and tertiary structures that RNA molecules can adopt.

Many approaches are available for investigating RNA structure. In addition to biophysical and biochemical techniques, computer modeling of RNA structure has become an important tool. Modeling programs can consider different types of information. For example, the known characteristics of RNA secondary structure, such as the ability to form double-stranded regions, can provide parameters for use in a modeling program.

A comparative approach can also be used in RNA structure prediction. This method assumes that RNAs of similar function and sequence have a similar structure. For example, the genes that encode certain types of RNAs, such as the 16S rRNA that makes up most of the small ribosomal subunit, have been sequenced from many different species. Among different species, the 16S rRNAs have similar but not identical sequences. Computer programs can compare many different 16S rRNA sequences to aid in the prediction of secondary structure. **Figure 22.7** illustrates a secondary structural model for 16S rRNA from *E. coli* based on a comparative sequence analysis of many bacterial 16S rRNA sequences. This large RNA contains 45 stem-loops. As you can imagine, it would be rather difficult to deduce such a model without the aid of a computer! In addition, RNA secondary structure prediction may be aimed at predicting the lowest energy state of a folded molecule. This approach, called free energy minimization, is also related to the base sequence of an RNA molecule.

Structure prediction is also used in the area of proteomics. As described in Chapter 15, proteins contain repeating secondary structural patterns known as α helices and β sheets. Several computer-based approaches attempt to predict secondary structure from the primary amino acid sequence. These programs base their predictions on different types of parameters. Some programs rely on the physical and energetic properties of the amino acids and the polypeptide backbone. More commonly, however, secondary structure predictions are based on the statistical frequency of amino acids within secondary structures that have already been crystallized.

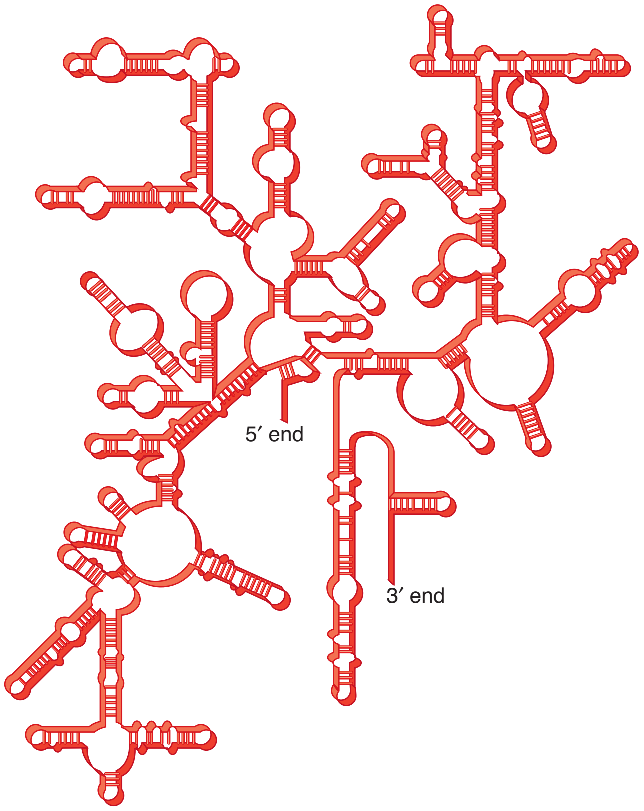


FIGURE 22.7 **A secondary structural model for *E. coli* 16S rRNA.**

Concept check: Would you expect the secondary structure of 16S rRNA from two closely related bacterial species to be similar? Explain.

For example, Peter Chou and Gerald Fasman compiled X-ray crystallographic data to calculate the likelihood that an amino acid will be found in an α helix or a β sheet. Certain amino acids, such as glutamic acid and alanine, are likely to be found in an α helix; others, such as valine and isoleucine, are more likely to be found in a β-sheet structure. Such information can be used to predict whether a sequence of amino acids within a protein is likely to be folded in an α-helix or β-sheet conformation. Secondary structure prediction is correct for approximately 60% to 70% of all sequences. Although this degree of accuracy is promising, it is generally not sufficient to predict protein secondary structure reliably. Therefore, one must be cautious in interpreting the results of a secondary structure prediction program.

Although the three-dimensional structure of a protein is extremely difficult to predict solely from its amino acid sequence, researchers have had some success in predicting tertiary structure using a comparative approach. This strategy requires that the protein of interest be homologous to another protein, the tertiary structure of which already has been solved by X-ray crystallography. In this situation, the crystal structure of the known protein can be used as a starting point to model the three-dimensional structure of the protein of interest. This approach is known as **homology-based modeling,** knowledge-based modeling, or comparative homology. As an example, two research groups independently predicted the three-dimensional structure of a protein, now known as HIV protease, that is encoded by a gene in HIV. The protein is homologous to other proteases with structures that had been solved by X-ray crystallography. Two similar models of the HIV protease were predicted before its actual crystal structure was determined. Both models turned out to be fairly accurate representations of the actual structure, which was later solved by X-ray crystallography. Knowledge of the structure of HIV protease helped scientists synthesize compounds known as HIV protease inhibitors, which are an important part of drug therapy to treat HIV/AIDS.

22.3 REVIEWING THE KEY CONCEPTS

- Bioinformatics involves the use of computers, mathematical tools, and statistical techniques to record, store, and analyze biological information, such as DNA sequences.
- Sequence files are analyzed by computer programs.
- Researchers have collected genetic sequences and compiled them in large databases (see Table 22.3).
- Different computational strategies, such as sequence recognition and pattern recognition, can be used to identify functional genetic sequences. Sequence recognition may identify sequence elements or motifs (see Table 22.4).
- Homologous genes are derived from the same ancestral gene. They can be orthologs (genes in different species) or paralogs (genes in the same species) (see Figure 22.6).
- The BLAST program is used to identify homologous sequences that are found within a database (see Table 22.5).
- Computer programs may also use genetic sequences to predict the structure of DNA, RNA, or proteins (see Figure 22.7).

22.3 COMPREHENSION QUESTIONS

1. The identification of a stop codon for a particular gene is an example of
   a. sequence recognition.
   b. pattern recognition.
   c. both a and b.
   d. none of the above.
2. Homologous genes
   a. are derived from the same ancestral gene.
   b. are likely to carry out the same or similar functions.
   c. have similar DNA sequences.
   d. exhibit all of the above.
3. The BLAST program begins with a particular genetic sequence and
   a. translates it into an amino acid sequence.
   b. determines if it contains one or more genes.
   c. identifies homologs within a database.
   d. does all of the above.

KEY TERMS

Page 531. functional genomics, proteome, proteomics, bioinformatics
Page 532. DNA microarray, gene chip, chromatin immunoprecipitation (ChIP)
Page 535. ChIP-chip assay, gene knockout
Page 536. alternative splicing
Page 537. RNA editing, posttranslational covalent modification, two-dimensional gel electrophoresis
Page 538. isoelectric focusing, mass spectrometry, tandem mass spectrometry
Page 539. protein microarrays, antibody microarray, functional protein microarray
Page 540. computer programs, computer data file
Page 541. database, annotated
Page 542. genome databases
Page 543. sequence recognition, pattern recognition, sequence element, motif, homologous, orthologs
Page 544. paralogs, gene family, homology, similarity, BLAST (basic local alignment search tool)
Page 545. E-value
Page 547. homology-based modeling

CHAPTER SUMMARY

22.1 Functional Genomics

- The goal of functional genomics is to understand the role of genetic sequences in a given species.
- A microarray is a slide dotted with many DNA sequences. It can be used to study the expression of many genes simultaneously and also has other uses (see Figure 22.1, Table 22.1).
- In a ChIP-chip assay, chromatin immunoprecipitation is used in conjunction with a microarray to study DNA-protein interactions at the genomic level (see Figure 22.2).
- Researchers are producing gene knockout collections for certain species such as mice to determine the functions of genes at the genomic level.

22.2 Proteomics

- The proteome is the entire collection of proteins that a cell or organism makes. The study of the function and interactions of many proteins is called proteomics.
- The proteome is much larger than the genome due to alternative splicing, RNA editing, and posttranslational covalent modifications (see Figure 22.3).
- Two-dimensional gel electrophoresis is used to separate a complex mixture of proteins (see Figure 22.4).
- Mass spectrometry and tandem mass spectrometry are used to identify short amino acids sequences within purified proteins.

These short sequences can be used to identify a protein (see Figure 22.5).
- Protein microarrays are used to study protein expression, protein function, protein-protein interactions, and protein-drug interactions (see Table 22.2).

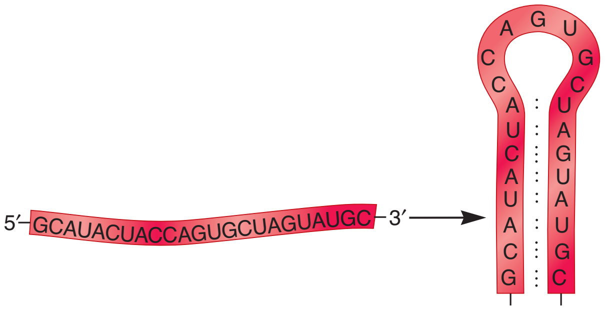
22.3 Bioinformatics

- Bioinformatics involves the use of computers, mathematical tools, and statistical techniques to record, store, and analyze biological information, such as DNA sequences.
- Sequence files are analyzed by computer programs.
- Researchers have collected genetic sequences and compiled them in large databases (see Table 22.3).
- Different computational strategies, such as sequence recognition and pattern recognition, can be used to identify functional genetic sequences. Sequence recognition may identify sequence elements or motifs (see Table 22.4).
- Homologous genes are derived from the same ancestral gene. They can be orthologs (genes in different species) or paralogs (genes in the same species) (see Figure 22.6).
- The BLAST program is used to identify homologous sequences that are found within a database (see Table 22.5).
- Computer programs may also use genetic sequences to predict the structure of DNA, RNA, or proteins (see Figure 22.7).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Using a comparative sequence analysis, the secondary structures of rRNAs have been predicted. Among many homologous rRNAs, one stem-loop usually has the following structure:


5′–GCAUACUACCAGUGCUAGUAUGC–3′
G C A U A C C A G U G C U A G U A U G C

You have sequenced the DNA that encodes a homologous rRNA from a new species and have obtained most of its sequence, but you cannot read the last five bases on your sequencing gel.

5′–GCATTCTACCAGTGCTAG?????–3′

Of course, you will eventually repeat this experiment to determine the last five bases. However, before you get around to doing this, what do you expect will be the sequence of the last five bases?

Answer: AATGC–3′

This will also form a similar stem-loop structure.

S2. Which of the following statements uses the term *homologous* correctly?

A. The two X chromosomes in female mammalian cells are homologous to each other.

B. The α-tubulin gene in *Saccharomyces cerevisiae* is homologous to the α-tubulin gene in *Arabidopsis thaliana*.

C. The promoter of the *lac* operon is homologous to the promoter of the *trp* operon.

D. The *lacY* gene of *E. coli* and *Klebsiella pneumoniae* are approximately 60% homologous to each other.

Answer:

A. Correct.

B. Correct.

C. Incorrect; they are short sequences that are similar to each other.

D. Incorrect; they are simply homologous. Two sequences are either homologous (i.e., derived from a common ancestral gene) or not. However, you could say the sequences are 60% identical to each other.

S3. The goal of many computer programs is to identify sequence elements within a long segment of DNA. What is a sequence element? Give two examples. How is the specific sequence of a sequence element determined? In other words, is it determined by the computer program or by genetic studies? Explain.

Answer:

A sequence element is a particular gene sequence that has some type of known functional role. One example is a promoter sequence that is needed for transcription. Another example is a start codon that is needed to begin the process of translation. Sequence elements are determined by experimentation.

Conceptual Questions

C1. Discuss the meaning of the following terms: *genomics, functional genomics,* and *proteomics.*

C2. Discuss the reasons why the proteome is larger than the genome of a given species.

C3. What is a database? What types of information are stored within a database? Where does the information come from? Discuss the objectives of a genome database.

C4. Besides the examples listed in Table 22.4, list five types of short sequences that a geneticist might want to locate within a DNA sequence.

C5. Discuss the distinction between sequence recognition and pattern recognition.

C6. What is the difference between similarity and homology?

Application and Experimental Questions

E1. With regard to DNA microarrays, answer the following questions:

A. What is attached to the slide? Be specific about the number of spots, the lengths of DNA fragments, and the origin of the DNA fragments.

B. What is hybridized to the microarray?

C. How is hybridization detected?

E2. For two-dimensional gel electrophoresis, what physical properties of proteins promote their separation in the first dimension and the second dimension?

E3. Can two-dimensional gel electrophoresis be used as a purification technique? Explain.

E4. Explain how tandem mass spectroscopy can be used to determine the sequence of a peptide. Once a peptide sequence is known, how is this information used to determine the sequence of the entire protein?

E5. Describe the two general types of protein microarrays. What are their possible applications?

E6. What is a motif? Why is it useful for computer programs to identify functional motifs within amino acid sequences?

E7. Discuss why it is useful to search a database to identify sequences that are homologous to a newly determined sequence.

E8. The secondary structure of 16S rRNA has been predicted using a computer-based sequence analysis. In general terms, discuss what type of information is used in a comparative sequence analysis, and explain what assumptions are made concerning the structure of homologous RNAs.

E9. Discuss the basis for secondary structure prediction in proteins. How reliable is it?

E10. To reliably predict the tertiary structure of a protein based on its amino acid sequence, what type of information must be available?

E11. In this chapter, we considered a computer program that can translate a DNA sequence into a polypeptide sequence. A researcher has a sequence file that contains the amino acid sequence of a polypeptide and runs a program that is opposite to this program. This other program is called BACKTRANSLATE. It can take an amino acid sequence file and determine the sequence of DNA that would encode such a polypeptide. How does this program work? In other words, what are the genetic principles that underlie this program? What type of sequence file would this program generate: a nucleotide sequence or an amino acid sequence? Would the BACKTRANSLATE program produce only a single sequence file? Explain why or why not.

E12. In this chapter, we considered a computer program that can translate a DNA sequence into a polypeptide sequence. Instead of running this program, a researcher could simply look the codons up in a genetic code table and determine the sequence by hand. What are the advantages of running this program rather than doing it the old-fashioned way by hand?

E13. To identify the following types of genetic occurrences, would a program use sequence recognition, pattern recognition, or both?

A. Whether a segment of *Drosophila* DNA contains a P element (which is a specific type of transposable element)

B. Whether a segment of DNA contains a stop codon

C. In a comparison of two DNA segments, whether there is an inversion in one segment compared with the other segment

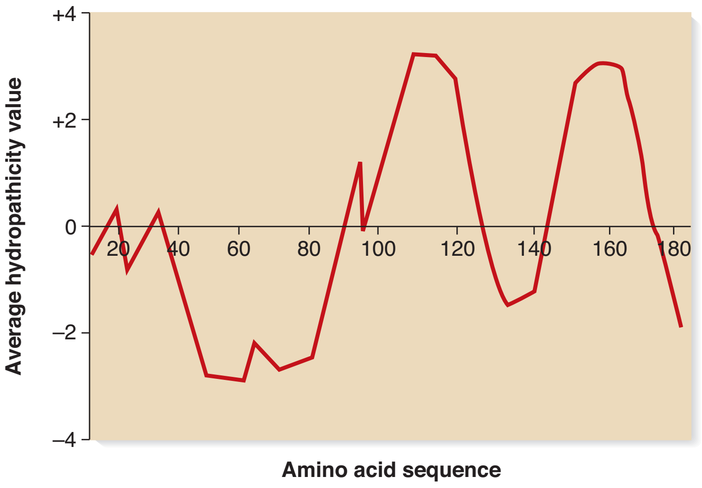
D. Whether a long segment of bacterial DNA contains one or more genes

E14. Here is a short nucleotide sequence within a gene. Via the Internet (e.g., see www.ncbi.nlm.nih.gov/Tools), determine what gene this sequence is found within. Also, determine the species in which this gene sequence is found.

5′–GGGCGCAATTACTTAACGCCTCGATTATCTTCTTGC
GCCACTGATCATTA–3′

E15. Membrane proteins often have transmembrane regions that span the membrane in an α-helical conformation. These transmembrane segments are about 20 amino acids long and usually contain amino acids with nonpolar (i.e., hydrophobic) amino acid side chains. Researchers can predict whether a polypeptide sequence has transmembrane segments based on the occurrence of segments that contain 20 nonpolar amino acids. To do so, each amino acid is assigned a hydropathy value, based on the chemistry of its amino acid side chain. Amino acids with very nonpolar side chains are given a high (positive) value, whereas amino acids that are charged and/or polar are given low (negative) values. The hydropathy values usually range from about +4 to −4.

Computer programs have been devised that scan the amino acid sequence of a polypeptide and calculate values based on the hydropathy values of the amino acid side chains. The program usually scans a window of seven amino acids and assigns an average hydropathy value. For example, the program would scan amino acids 1 through 7 and give an average value, then it would scan 2 through 8 and give a value, then it would scan 3 through 9 and give a value, and so on, until it reached the end of the polypeptide sequence (i.e., until it reached the carboxyl terminus). The program then produces a figure, known as a hydropathy plot, which describes the average hydropathy values throughout the entire polypeptide sequence. An example of a hydropathy plot is shown here.



A. How many transmembrane segments are likely in this polypeptide?

B. Draw the structure of this polypeptide if it were embedded in the plasma membrane. Assume that the amino terminus is found in the cytoplasm of the cell.

E16. Explain how a computer program can predict RNA secondary structure. What is the underlying genetic concept used by the program to predict secondary structure?

E17. Are the following statements about protein structure prediction true or false?

A. The prediction of secondary structure relies on information regarding the known occurrence of amino acid residues in α helices or β sheets from X-ray crystallographic data.

B. The prediction of secondary structure is highly accurate, nearly 100% correct.

C. To predict the tertiary structure of a protein based on its amino acid sequence, it is necessary that the protein of interest is homologous to another protein whose tertiary structure is already known.

Questions for Student Discussion/Collaboration

1. Let's suppose you are in charge of organizing and publicizing a genomic database for the mouse genome. Make a list of innovative strategies you would initiate to make the mouse genome database useful and effective.
2. Let's suppose a 5-year-old told you that she was interested in pursuing a career studying the three-dimensional structure of proteins. (Okay, so she's a bit precocious.) Would you advise her to become a geneticist, a mathematical theoretician, or a biophysicist?
3. If you have access to the necessary computer software, make a sequence file and analyze it in the following ways: What is the translated sequence in all three reading frames? What is the longest open reading frame? Is the sequence homologous to any known sequences? If so, does this provide any clues about the function of the sequence?

Answers to Comprehension Questions

22.1: c, b, c, a

22.2: d, d, a, d

22.3: a, d, c

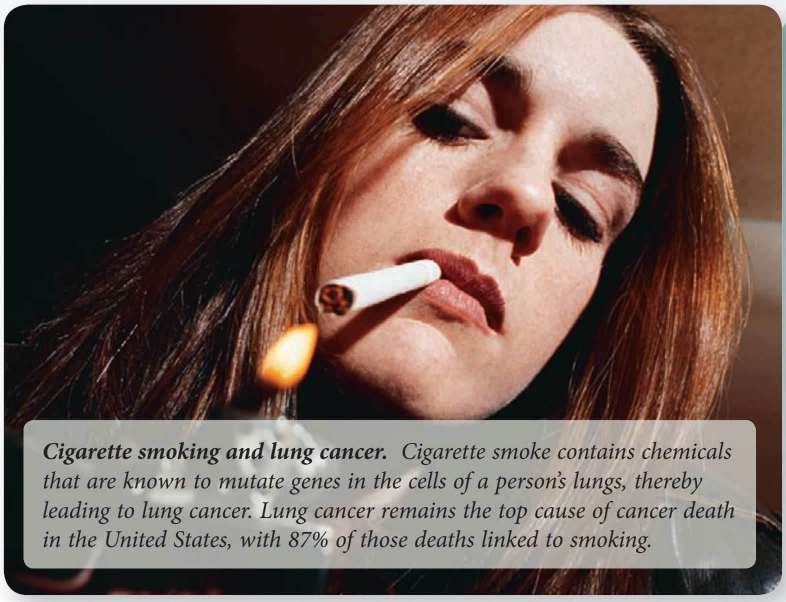
Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

PART VI GENETIC ANALYSIS OF INDIVIDUALS AND POPULATIONS

23



Cigarette smoking and lung cancer. *Cigarette smoke contains chemicals that are known to mutate genes in the cells of a person's lungs, thereby leading to lung cancer. Lung cancer remains the top cause of cancer death in the United States, with 87% of those deaths linked to smoking.*

CHAPTER OUTLINE

23.1 Inheritance Patterns of Genetic Diseases
23.2 Detection of Disease-Causing Alleles
23.3 Prions
23.4 Genetic Basis of Cancer

MEDICAL GENETICS AND CANCER

Genetic information is highly personal and unique. Our genes underlie every aspect of human health, both in function and dysfunction. Obtaining a detailed understanding of how genes work together and interact with environmental factors ultimately allows us to appreciate the differences between the events in normal cellular processes and those that occur in disease pathogenesis. Such knowledge profoundly affects the way many diseases are defined, diagnosed, treated, and prevented. Genetic insight is expected to bring about revolutionary changes in medical practices. In fact, changes have already begun. Currently, several hundred genetic tests are in clinical use, with many more under development. Most of these tests detect mutations associated with rare genetic disorders that follow Mendelian inheritance patterns. These include Duchenne muscular dystrophy, cystic fibrosis, sickle cell disease, and Huntington disease. In addition, genetic tests are available to detect the predisposition to develop certain forms of cancer.

Approximately 12,000 genetic diseases are known to afflict people, but this is almost certainly an underestimate. The website Online Mendelian Inheritance in Man (OMIM) is a compendium of human genes and phenotypes containing information on all known disorders that are inherited in a Mendelian manner (see www.ncbi.nlm.nih.gov/omim/).

Most of the genetic disorders discussed in the first part of this chapter are the direct result of a mutation in one gene. However, many diseases have a complex pattern of inheritance involving several genes. These include common medical disorders such as diabetes, asthma, and mental illness. In these cases, a single mutant gene does not determine whether a person has a disease. Instead, a number of genes may each make a subtle contribution to a person's susceptibility to a disease. Unraveling these complexities will be a challenge for some time to come. The availability of the human genome sequence, discussed in Chapter 21, will be of great help.

In this chapter, we will focus our attention on ways that mutant genes contribute to human disease. In the first part of the chapter, we explore the molecular basis of several genetic disorders and their patterns of inheritance. We also examine how genetic testing can determine if an individual carries a defective allele. The last part of the chapter concerns cancer, a disease that involves the uncontrolled growth of somatic cells. We will examine the underlying genetic basis for cancer and discuss the roles that many different genes may play in the development of this disease.

23.1 INHERITANCE PATTERNS OF GENETIC DISEASES

Human genetics is a topic that is hard to resist. Almost everyone who looks at a newborn is tempted to speculate whether the baby resembles the mother, the father, or perhaps a distant relative. In this section, we will focus primarily on the inheritance of human genetic diseases rather than common traits found in the general population. Even so, the study of human genetic diseases provides insights regarding our traits. The disease hemophilia illustrates this point. Hemophilia is a condition in which the blood does not clot properly. By analyzing people with this disorder, researchers have identified genes that participate in the process of blood clotting. The study of hemophilia has helped to elucidate a clotting pathway involving several different proteins. Therefore, as with the study of mutants in model organisms such as *Drosophila,* mice, and yeast, when we study the inheritance of genetic diseases, we often learn a great deal about the genetic basis for normal physiological processes as well.

Because thousands of human diseases have an underlying genetic basis, human genetic analysis is of great medical importance. In this section, we will examine the causes and inheritance patterns of human genetic diseases that result from defects in single genes. As you will learn, the mutant genes that cause these diseases often follow simple Mendelian inheritance patterns.

A Genetic Basis for a Human Disease May Be Suggested from a Variety of Observations

When we view the characteristics of people, we usually think that some traits are inherited, whereas others are caused by environmental factors. For example, when the facial features of two related individuals look strikingly similar, we think that this similarity has a genetic basis. The profound resemblance between identical twins is an obvious example. By comparison, other traits are governed by the environment. If we see a person with purple hair, we likely suspect that he or she has used hair dye as opposed to showing an unusual genetic trait.

For human diseases, geneticists would like to know the relative contributions from genetics and the environment. Is a disease caused by a pathogenic microorganism, a toxic agent in the environment, or a faulty gene? Unlike the case with experimental organisms, we cannot conduct human crosses to determine the genetic basis for diseases. Instead, we must rely on analyzing the occurrence of a disease in families that already exist. As described in the following list, several observations are consistent with the idea that a disease is caused, at least in part, by the inheritance of mutant genes. When the occurrence of a disease correlates with several of these observations, a geneticist becomes increasingly confident that it has a genetic basis.

1. *When an individual exhibits a disease, this disorder is more likely to occur in genetic relatives than in the general population.* For example, someone with cystic fibrosis (CF) is more likely to have relatives with this disease than would a randomly chosen member of the general population.
2. *Identical twins share the disease more often than nonidentical twins.* Identical twins, also called **monozygotic (MZ) twins,** are genetically identical to each other, because they were formed from the same sperm and egg. By comparison, nonidentical twins, also called fraternal, or **dizygotic (DZ) twins,** are formed from separate pairs of sperm and egg cells. Fraternal twins share, on average, 50% of their genetic material, the same as any two siblings. When a disorder has a genetic component, identical twins are more likely to exhibit the disorder than are fraternal twins.

   Geneticists evaluate a disorder's **concordance,** the degree to which it is inherited, by calculating the percentage of twin pairs in which both twins exhibit the disorder relative to pairs in which only one twin shows the disorder. Theoretically, for diseases caused by a single gene, concordance among identical twins should be 100%. For dominant disorders, concordance among fraternal twins is expected to be 50%, assuming only one parent is heterozygous for the disease. For recessive diseases, concordance among fraternal twins would be 25% if we assume both parents are heterozygous carriers. However, the actual concordance values observed for most single-gene disorders are usually less than such theoretical values for a variety of reasons. Some disorders are not completely penetrant, meaning that the symptoms associated with the disorder are not always produced. Also, one twin may have a disorder due to a new mutation that occurred after fertilization. It is very unlikely that the other twin would have the same mutation.
3. *The disease does not spread to individuals sharing similar environmental situations.* Inherited disorders cannot spread from person to person. The only way genetic diseases can be transmitted is from parent to offspring.
4. *Different populations tend to have different frequencies of the disease.* Due to evolutionary factors, the frequencies of traits usually vary among different populations of humans. For example, the frequency of sickle cell disease is highest among certain African and Asian populations and relatively low in other parts of the world (see Chapter 25, Figure 25.12).
5. *The disease tends to develop at a characteristic age.* Many genetic disorders exhibit a characteristic **age of onset** at which the disease appears. Some mutant genes exert their effects during embryonic and fetal development, so their effects are apparent at birth. Other genetic disorders tend to develop much later in life.
6. *The human disorder may resemble a disorder that is already known to have a genetic basis in an animal.* In animals, where we can conduct experiments, various traits are known to be governed by genes. For example, the albino phenotype is found in humans as well as in many animals (**Figure 23.1**).

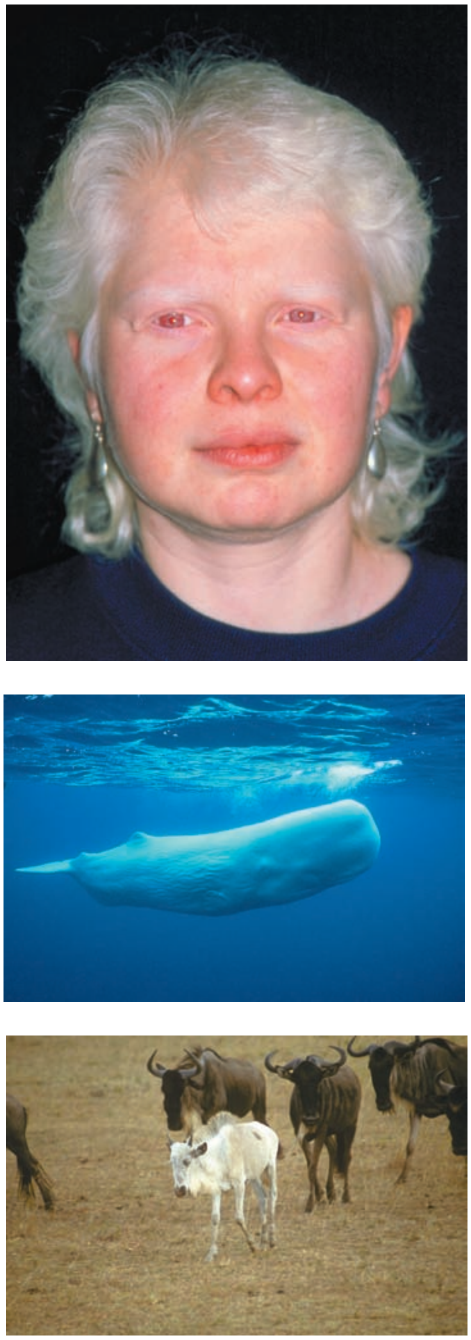


FIGURE 23.1 **The albino phenotype in humans and other mammals.**

Genes→Traits Certain enzymes (encoded by genes) are necessary for the production of pigment. A homozygote with two defective alleles in these pigmentation genes exhibits an albino phenotype. This phenotype can occur in humans and other animals.

7. *A correlation is observed between a disease and a mutant human gene or a chromosomal alteration.* A particularly convincing piece of evidence that a disease has a genetic basis is the identification of altered genes or chromosomes that occur only in people exhibiting the disorder. When comparing two individuals, one with a disease and one without, we expect to see differences in their genetic material if the disorder has a genetic component. Alterations in gene sequences can be determined by gene cloning and DNA-sequencing techniques (see Chapter 19). Also, changes in chromosome structure and number can be detected by the microscopic examination of chromosomes (see Chapter 8).

Inheritance Patterns of Human Diseases May Be Determined via Pedigree Analysis

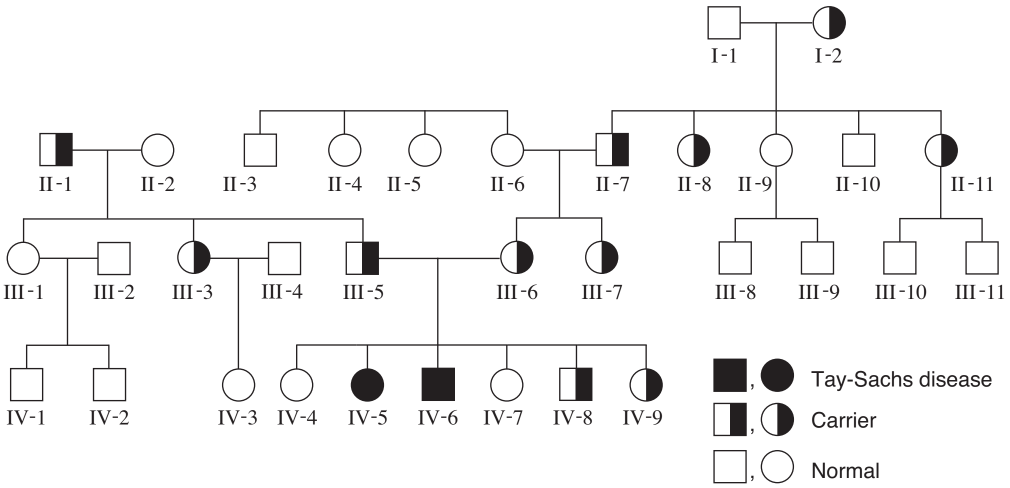
When a human disorder is caused by a mutation in a single gene, the pattern of inheritance can be deduced by analyzing human pedigrees. How is this accomplished? A geneticist must obtain data from many large pedigrees containing several individuals who exhibit the disorder and then follow its pattern of inheritance from generation to generation. The basic features of pedigrees are discussed in Chapter 3. You may wish to review Figure 3.13 for the organization and symbols of pedigrees. In this chapter, we will examine a few large pedigrees involving human diseases that are inherited in different ways.

Autosomal Recessive Inheritance The pedigree shown in Figure 23.2 concerns a genetic disorder called Tay-Sachs disease (TSD), first described by Warren Tay, a British ophthalmologist, and Bernard Sachs, an American neurologist, in the 1880s. Affected individuals appear healthy at birth but then develop neurodegenerative symptoms at 4 to 6 months of age. The primary characteristics are cerebral degeneration, blindness, and loss of motor function. Individuals with TSD typically die in the third or fourth year of life. This disease is particularly prevalent in Ashkenazi (eastern European) Jewish populations, in which it has a frequency of about 1 in 3600 births, which is over 100 times more frequent than in most other human populations.

At the molecular level, the mutation that causes TSD is in a gene that encodes the enzyme hexosaminidase A (HexA). HexA is responsible for the breakdown of a category of lipids called G_{M2}-gangliosides, which are prevalent in the cells of the central nervous system. A defect in the ability to break down this lipid leads to its excessive accumulation in nerve cells and eventually causes the neurodegenerative symptoms characteristic of TSD. This defect in lipid breakdown was recognized long before the *hexA* gene was identified as being defective in these patients.

As illustrated in Figure 23.2, Tay-Sachs disease is inherited in an autosomal recessive manner. The four common features of autosomal recessive inheritance are as follows:

1. *An affected offspring may have two unaffected parents.* For rare recessive traits, the parents are usually unaffected, meaning they do not exhibit the disease. For deleterious alleles that cause early death or infertility, the two parents must be unaffected. This is always the case in TSD.
2. *When two unaffected heterozygotes have children, the percentage of affected children is (on average) 25%.*
3. *Two affected individuals have 100% affected children.* This observation can be made only when a recessive trait produces fertile, viable individuals. In the case of TSD, the



INTERACTIVE EXERCISE

FIGURE 23.2 **A family pedigree of Tay-Sachs disease, indicating recessive inheritance.** Heterozygous carriers were determined by genetic testing.

Concept check: *What feature(s) of this pedigree indicate recessive inheritance?*

affected individual dies in early childhood, and so it is not possible to observe crosses between two affected people.

4. *The trait occurs with the same frequency in both sexes.*

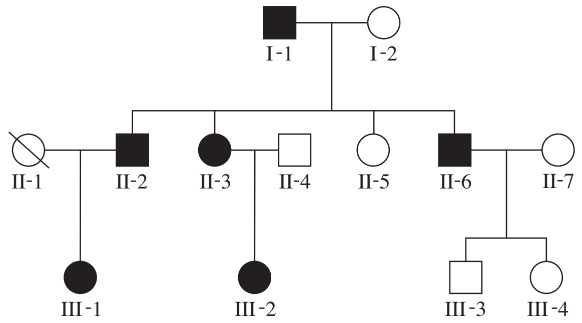
Autosomal recessive inheritance is a common mode of transmission for genetic disorders, particularly those that involve defective enzymes. Human recessive alleles are often the result of **loss-of-function mutations** that cause a reduction or loss of function in the encoded protein. In the case of Tay-Sachs disease, the functional protein is the enzyme hexosaminidase A. Heterozygous carriers have approximately 50% of the functional enzyme, which is sufficient for a normal phenotype. However, for other genetic diseases, the level of functional protein may vary due to effects of gene regulation. Thousands of human genetic diseases are inherited in a recessive manner, and in many cases, the mutant genes have been identified. A few of these diseases are described in **Table 23.1.**

Autosomal Dominant Inheritance Now let's examine a human pedigree involving an autosomal dominant disease (**Figure 23.3**). In this example, the affected individuals have a disorder called Huntington disease (also called Huntington chorea). The major symptoms of this disease, which usually occurs during middle age, are due to the degeneration of certain types of neurons in the brain, leading to personality changes, dementia, and early death. There is no cure for the disease. In 1993, the gene involved in Huntington disease was identified and sequenced. It encodes a protein called huntingtin that is expressed in neurons but is also found in some cells not affected in Huntington disease. In persons with this disorder, a mutation called a trinucleotide repeat expansion produces a very long polyglutamine tract—many glutamines in a row—within the huntingtin protein (see Chapter 18). This causes an aggregation of the protein in

TABLE **23.1**

Examples of Human Disorders Inherited in an Autosomal Recessive Manner

Disorder	Chromosomal Location of Gene	Gene Product	Effective of Disease-Causing Allele
Adenosine deaminase deficiency	20q	Adenosine deaminase	Defective immune system and skeletal and neurological abnormalities
Albinism (type I)	11q	Tyrosinase	Inability to synthesize melanin, resulting in white skin, hair, etc.
Cystic fibrosis (CF)	7q	CF transmembrane conductance regulator	Water imbalance in tissues of the pancreas, intestine, sweat glands, and lungs due to impaired ion transport; leads to lung damage
Phenylketonuria (PKU)	12q	Phenylalanine hydroxylase	Foul-smelling urine, neurological abnormalities, mental impairment; may be remedied by diet modification at birth
Sickle cell disease	11p	β globin	Anemia, blockages in blood circulation
Tay-Sachs disease (TSD)	15q	Hexosaminidase A	Progressive neurodegeneration



INTERACTIVE EXERCISE

FIGURE 23.3 **A family pedigree of Huntington disease, indicating dominant inheritance.**

Concept check: *What feature(s) of this pedigree indicate dominant inheritance?*

neurons. However, additional research is needed to elucidate the molecular relationship between the abnormality in the huntingtin protein and the disease symptoms.

Five common features of autosomal dominant inheritance are:

1. *An affected offspring usually has one or both affected parents.* However, this is not always the case. Some dominant traits show incomplete penetrance (see Chapter 5), so a heterozygote may not exhibit the trait even though it may be passed to offspring who do exhibit the trait. Also, a dominant mutation may occur during gametogenesis, so two unaffected parents may produce an affected offspring.
2. *An affected individual with only one affected parent is expected to produce 50% affected offspring (on average).*
3. *Two affected, heterozygous individuals have (on average) 25% unaffected offspring.*
4. *The trait occurs with the same frequency in both sexes.*
5. *For most dominant, disease-causing alleles, the homozygote is more severely affected with the disorder. In some cases, a dominant allele may be lethal in the homozygous condition.*

Numerous autosomal dominant diseases have been identified in humans (Table 23.2). The three common explanations for dominant disorders are haploinsufficiency, a gain-of-function mutation, or a dominant-negative mutation. Let's consider examples of all three types.

The term **haploinsufficiency** refers to the phenomenon in which a person has only a single functional copy of a gene, and that single functional copy does not result in a normal phenotype. In these disorders, 50% of the functional protein is not sufficient to produce a normal phenotype. Haploinsufficiency shows a dominant pattern of inheritance because a heterozygote (with one normal allele and one inactive allele) has the disease. An example is aniridia, which is a rare disorder that results in an absence of the iris of the eye. Aniridia leads to visual impairment and blindness in severe cases.

A second category of dominant disorders involves **gain-of-function mutations.** Such mutations change the gene product so it "gains" a new and abnormal function. An example is achondroplasia, which is characterized by abnormal bone growth that results in short stature with relatively short arms and legs. This disorder is caused by a mutation that occurs in a gene that encodes the fibroblast growth factor receptor-3, which has a negative effect on bone growth. In achondroplasia, the mutant form of the receptor is overactive. This hyperactivity causes bone growth to be inhibited prematurely, thereby leading to severely shortened bones.

Finally, some dominant disorders are caused by **dominant-negative mutations** in which the altered gene product acts

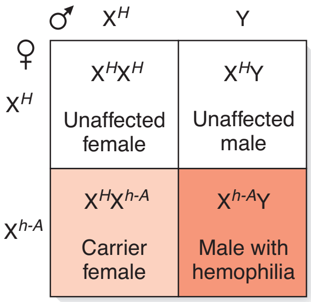
TABLE **23.2**

Examples of Human Disorders Inherited in an Autosomal Dominant Manner

Disorder	Chromosomal Location of Gene	Gene Product	Effects of Disease-Causing Allele
Aniridia	11p	Pax6 transcription factor	An absence of the iris of the eye, leading to visual impairment and sometimes blindness
Achondroplasia	4p	Fibroblast growth factor receptor-3	A common form of dwarfism associated with a defect in the growth of long bones
Marfan syndrome	15q	Fibrillin-1	Tall and thin individuals with abnormalities in the skeletal, ocular, and cardiovascular systems due to a weakening in the elasticity of certain body parts
Osteoporosis	7q	Collagen (type $1_{\alpha 2}$)	Brittle, weakened bones
Familial hypercholesterolemia	19p	LDL receptor	Very high serum levels of low-density lipoprotein (LDL), a predisposing factor in heart disease
Huntington disease	4p	Huntingtin	Neurodegeneration that occurs relatively late in life, usually in middle age
Neurofibromatosis I	17q	Neurofibromin	Individuals may exhibit spots of abnormal pigmentation (café-au-lait spots) and growth of noncancerous tumors in the nervous system

antagonistically to the normal gene product. In humans, Marfan syndrome, which is due to a mutation in the *fibrillin-1* gene, is an example. The *fibrillin-1* gene encodes a glycoprotein that is a structural component of the extracellular matrix that provides structure and elasticity to tissues. The mutant gene encodes a fibrillin-1 protein that antagonizes the effects of the normal protein, thereby weakening the elasticity of certain body parts. For example, the walls of the major arteries such as the aorta, the large artery that leaves the heart, are often affected.

X-Linked Recessive Inheritance Let's now turn to another inheritance pattern common in humans that is called X-linked recessive inheritance (**Table 23.3**). Recessive X-linked inherited diseases pose a special problem for males. Why are males more likely to be affected? Most X-linked genes lack a counterpart on the Y chromosome. Males are hemizygous—have a single copy—for these genes. Therefore, a female heterozygous for an X-linked recessive gene passes this trait on to 50% of her sons, as shown in the following Punnett square for hemophilia. In this example, $X^{h\text{-}A}$ is the chromosome that carries a mutant allele causing hemophilia, whereas X^H carries the wild-type allele.



♀	♂ X^H	Y
X^H	X^HX^H Unaffected female	X^HY Unaffected male
$X^{h\text{-}A}$	$X^HX^{h\text{-}A}$ Carrier female	$X^{h\text{-}A}Y$ Male with hemophilia

As mentioned previously, hemophilia is a disorder in which the blood cannot clot properly when a wound occurs. For individuals with this trait, a minor cut may bleed for a very long time, and small injuries can lead to large bruises, because internal broken capillaries may leak blood profusely before they are repaired. For hemophiliacs, common injuries pose a threat of severe internal or external bleeding. Hemophilia A, also called classical hemophilia, is caused by a defect in an X-linked gene that encodes the protein clotting factor VIII. This disease has also been called the "royal disease," because it affected many members of European royal families during the 19th and early 20th centuries. The pedigree shown in **Figure 23.4** illustrates the prevalence of hemophilia A among the descendants of Queen Victoria of England. The pattern of X-linked recessive inheritance is revealed by the following observations:

1. *Males are much more likely to exhibit the trait.*
2. *The mothers of affected males often have brothers or fathers who are affected with the same trait.*
3. *The daughters of affected males produce, on average, 50% affected sons.*

TABLE **23.3**

Examples of Human Disorders Inherited in an X-linked Recessive Manner

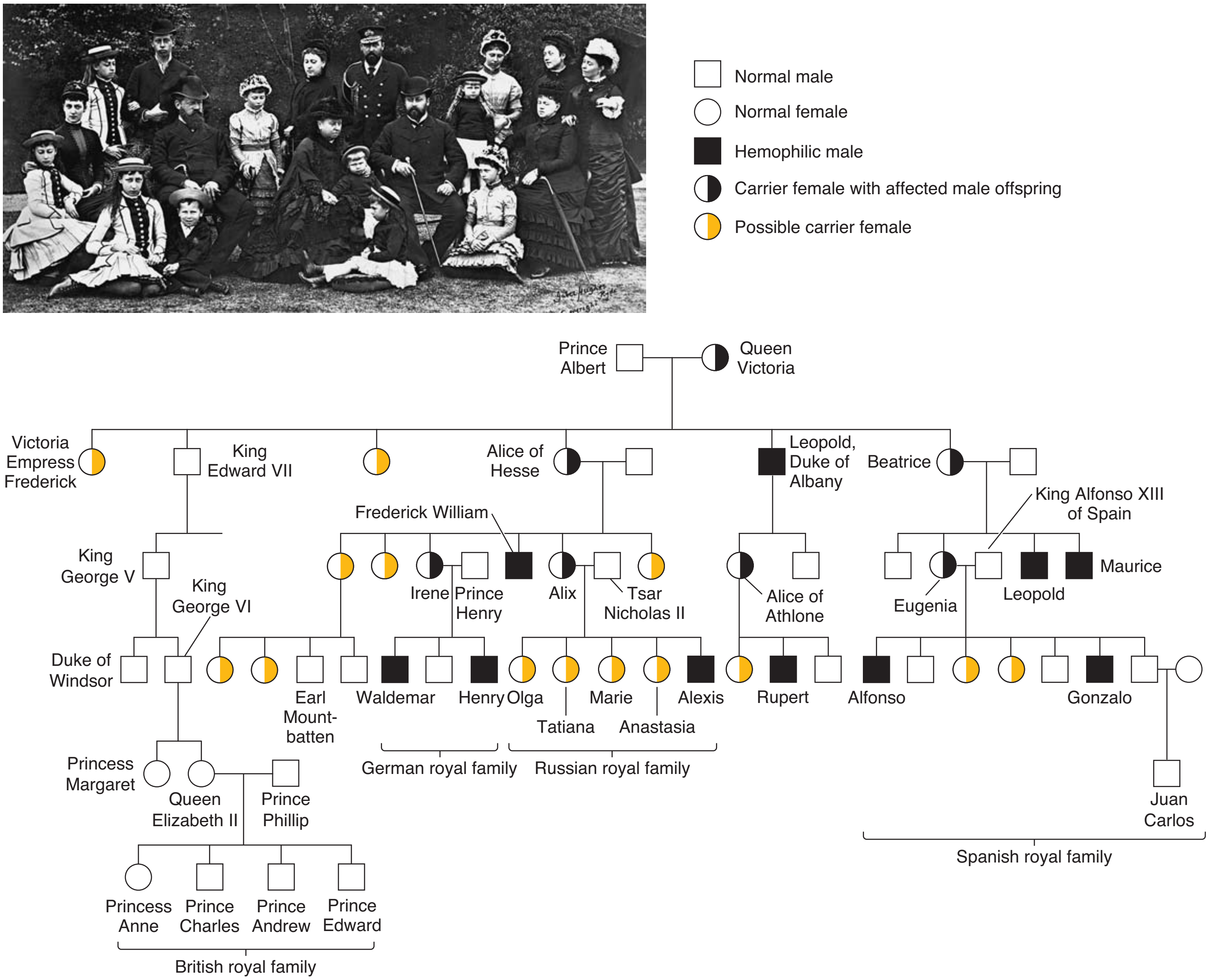
Disorder	Gene Product	Effects of Disease-Causing Allele
Duchenne muscular dystrophy	Dystrophin	Progressive degeneration of muscles that begins in early childhood
Hemophilia A	Clotting factor VIII	Defect in blood clotting
Hemophilia B	Clotting factor IX	Defect in blood clotting
Androgen insensitivity syndrome	Androgen receptor	Missing male steroid hormone receptor; XY individuals have external features that are feminine but internally have undescended testes and no uterus

X-Linked Dominant Inheritance Relatively few genetic disorders in humans follow an X-linked dominant inheritance pattern. In most cases, males are more severely affected than females, because females usually carry an X chromosome with a normal copy of the gene in question. In most of the X-linked dominant disorders listed in **Table 23.4**, male embryos die at an early stage of development, so most individuals exhibiting the disorder are females. Also, due to the dominant nature and severity of the disorders, persons with some of the disorders listed in Table 23.4 do not reproduce. Therefore, these disorders, which include Rett syndrome and Aicardi syndrome, are always caused by new mutations that occur during gamete formation or early embryogenesis. For those X-linked dominant disorders in which the offspring can reproduce, the following inheritance pattern is often observed:

1. *Females are much more likely to exhibit the trait when it is lethal to males.*
2. *Affected mothers have a 50% chance of passing the trait to daughters. Note: Affected mothers also have a 50% chance of passing the trait to sons, but for many of these disorders, affected sons are not observed because of lethality.*

Many Genetic Disorders Exhibit Locus Heterogeneity

Hemophilia can be used to illustrate another concept in genetics called **locus heterogeneity**—the phenomenon in which a particular type of disease may be caused by mutations in two or more different genes. For example, blood clotting involves the participation of several different proteins that take part in a cellular cascade that leads to the formation of a clot. Hemophilia is usually caused by a defect in one of three different clotting factors. In hemophilia A, also called classic hemophilia, a protein called factor VIII is missing. Hemophilia B is a deficiency in a different clotting factor, called factor IX. Both factor VIII and IX are encoded by different genes on the X chromosome. These two types of hemophilia show an X-linked recessive pattern of inheritance. By comparison, hemophilia C is due to a factor XI deficiency. The gene encoding factor XI is found on


Normal male
Normal female
Hemophilic male
Carrier female with affected male offspring
Possible carrier female
Prince Albert
Queen Victoria
Victoria Empress Frederick
King Edward VII
Alice of Hesse
Leopold, Duke of Albany
Beatrice
Frederick William
King Alfonso XIII of Spain
King George V
King George VI
Irene
Prince Henry
Alix
Tsar Nicholas II
Alice of Athlone
Eugenia
Leopold
Maurice
Duke of Windsor
Earl Mount-batten
Waldemar
Henry
Olga
Marie
Alexis
Rupert
Alfonso
Gonzalo
Tatiana
Anastasia
German royal family
Russian royal family
Princess Margaret
Queen Elizabeth II
Prince Phillip
Juan Carlos
Spanish royal family
Princess Anne
Prince Charles
Prince Andrew
Prince Edward
British royal family

INTERACTIVE EXERCISE

FIGURE 23.4 **A family tree of hemophilia A in the royal families of Europe, indicating an X-linked recessive inheritance.** Pictured are Queen Victoria and Prince Albert of Great Britain with some of their descendants.

Concept check: *What feature(s) of this pedigree indicate X-linked recessive inheritance?*

TABLE 23.4

Examples of Human Disorders Inherited in an X-linked Dominant Manner

Disorder	Gene Product	Effects of Disease-Causing Allele
Vitamin D-resistant rickets	Metallopeptidase	Defects in bone mineralization at the sites of bone growth or remodeling, leading to bone deformity and stunted growth in children
Rett syndrome	Methyl-CpG-binding protein-2	A neurodevelopmental disorder that includes a deceleration of head growth and small hands and feet; fatal in males.
Aicardi syndrome	Unknown	Characterized by the partial or complete absence of a key structure in the brain called the corpus callosum, and the presence of retinal abnormalities; fatal in males.
Incontinentia pigmenti	NFκβ essential modulator	Characterized by morphological and pigmentation abnormalities in the skin, hair, teeth, and nails; fatal in males.

chromosome 4, and this form of hemophilia follows an autosomal recessive pattern of inheritance, equally affecting males and females.

In hemophilia, locus heterogeneity arises from the participation of several proteins in a common cellular process. Another mechanism that may lead to locus heterogeneity occurs when proteins are composed of two or more different subunits, with each subunit being encoded by a different gene. The disease thalassemia is an example of locus heterogeneity caused by a mutation in a protein composed of multiple subunits. This potentially life-threatening disease involves defects in the ability of the red blood cells to transport oxygen. The underlying cause is an alteration in hemoglobin. In adults, hemoglobin is a tetrameric protein composed of two α-globin and two β-globin subunits; α globin and β globin are encoded by separate genes (namely, the α-globin and β-globin genes). Two main types of thalassemia have been discovered in human populations: α thalassemia, in which the α-globin subunit of hemoglobin is defective; and β thalassemia, in which the β-globin subunit is defective.

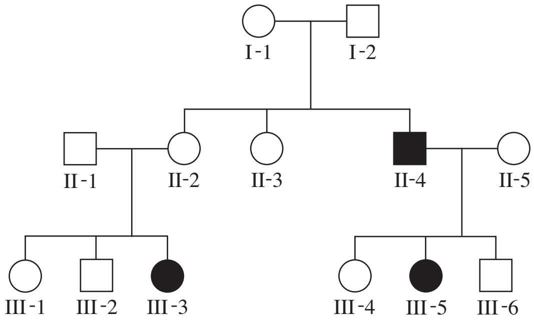
Unfortunately, locus heterogeneity may greatly confound pedigree analysis. For example, a human pedigree might contain some individuals with X-linked hemophilia and other individuals with hemophilia C. A geneticist who assumed all affected individuals had defects in the same gene would be unable to explain the resulting pattern of inheritance. For disorders such as hemophilia and thalassemia, pedigree analysis is not a major problem because the biochemical basis for these diseases is well understood. However, for rare diseases that are poorly understood at the molecular level, locus heterogeneity may profoundly obscure the pattern of inheritance.

23.1 REVIEWING THE KEY CONCEPTS

- A genetic basis for a human disease may be suggested by a variety of different observations (see Figure 23.1).
- Thousands of human genetic diseases follow a simple Mendelian pattern of inheritance. These include autosomal recessive, autosomal dominant, X-linked recessive, and X-linked dominant diseases (see Figures 23.2–23.4, Tables 23.1–23.4).
- Recessive diseases are usually caused by loss-of-function mutations, whereas dominant diseases may be caused by haploinsufficiency, gain-of-function mutations, or dominant-negative mutations.
- Many genetic diseases exhibit locus heterogeneity, which means that mutations in more than one gene can cause the disorder.

23.1 COMPREHENSION QUESTIONS

1. Which of the following would *not* be consistent with the idea that a disorder has a genetic component?
   a. The disorder is more likely to occur in a person's relatives.
   b. The disorder can spread to individuals sharing similar environments.
   c. The disorder tends to develop at a characteristic age.
   d. A correlation is observed between the disorder and a mutant gene.
2. Assuming complete penetrance, which type of inheritance pattern is consistent with the pedigree shown here?



   a. Autosomal recessive
   b. Autosomal dominant
   c. X-linked recessive
   d. X-linked dominant
3. Which of the following is *not* a common explanation for a dominant disorder?
   a. Haploinsufficiency
   b. A change in chromosome number
   c. A gain-of-function mutation
   d. A dominant-negative mutation
4. Locus heterogeneity refers to a genetic disorder that
   a. has a heterogeneous phenotype.
   b. is caused by mutations in two or more different genes.
   c. involves a structural change in multiple chromosomes.
   d. is inherited from both parents.

23.2 DETECTION OF DISEASE-CAUSING ALLELES

Because mutant genes are known to play a role in thousands of diseases, many insights regarding the disease process can be gained from understanding how a mutant gene causes a particular disease. The study of loss-of-function and gain-of-function alleles often provides important clues as to which cellular processes are altered in the disease state. Therefore, researchers have devoted great effort to identifying alleles associated with genetic diseases. In this section, we will explore various approaches used to identify mutant alleles that cause disease and also consider methods that can determine if an individual carries a disease-causing allele.

Haplotypes Exhibit Genetic Variation

To identify disease-causing alleles, researchers often rely on the known locations of genes and molecular markers along chromosomes that have been characterized in human populations. A disease-causing allele may be identified due to its proximity to another known gene or its proximity to molecular markers.

As discussed in Chapter 21, researchers can characterize chromosomes at the molecular level and determine the precise locations of genes and molecular markers along each chromosome. During the course of evolution, new mutations arise that alter the DNA sequences of genes and molecular markers. For this

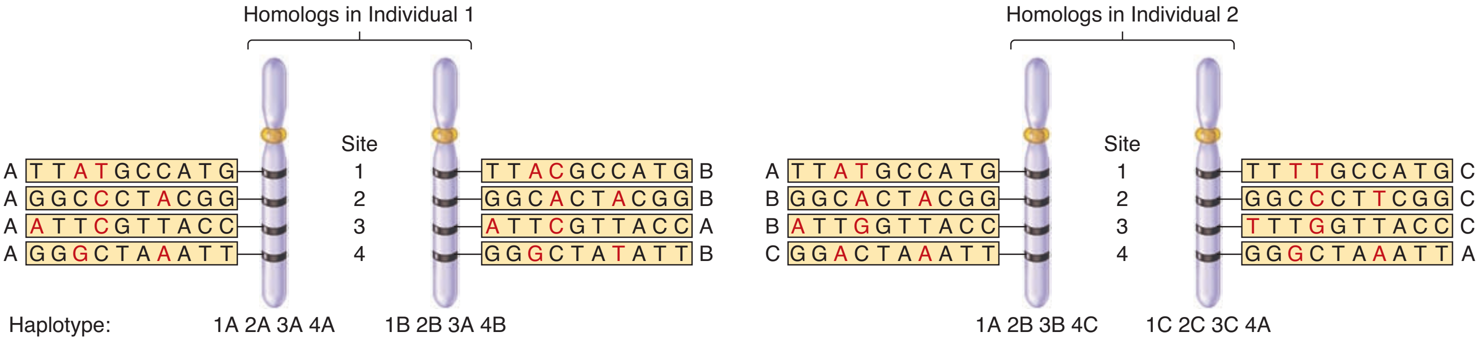


FIGURE 23.5 **A schematic representation of haplotypes along a human chromosome.** This example considers four sites along a chromosome that may exist in different versions designated A, B, or C. Bases that differ are shown in red.

Concept check: *What is a haplotype?*

reason, homologous chromosomes exhibit genetic differences (i.e., allelic variation) and show variation in their molecular markers.

As an example, **Figure 23.5** considers a pair of homologous chromosomes from two different individuals and focuses on four sites (called 1, 2, 3, and 4) that occur at particular locations along those chromosomes. These sites could be within particular genes or they could be molecular markers used in mapping studies. In this drawing, each site is also given a letter designation (A, B, or C) depending on the variation in the DNA sequence at the site. In individual 1, sites 1, 2, and 4 differ at one base pair between the homologs. In individual 2, all four sites differ at one base pair. Also note that individuals 1 and 2 differ with regard to some of these sites. These genetic differences at the four sites are examples of single-nucleotide polymorphisms, which are discussed in Chapter 25.

The term **haplotype,** which is a contraction for haploid genotype, refers to the linkage of alleles or molecular markers along a small region of a single chromosome. In Figure 23.5, the haplotype for these four sites are shown at the bottom of each chromosome. For example, the haplotype of the left homolog in individual 2 is 1A 2B 3B 4C.

Because mutations are rare events, haplotypes do not dramatically change from one generation to the next due to new mutations. By comparison, haplotypes are more likely to change over the course of a few generations due to crossing over. However, the likelihood of changing a haplotype depends on the distance between the sites. If two sites are far apart, a crossover is more likely to alter their pattern than if they are close together (see Chapter 7). If sites 1, 2, 3, and 4 were very close together along this chromosome, the haplotypes shown in this figure would be likely to stay the same after a few generations. For example, a great-great-great grandchild of individual 2 may inherit the haplotype 1A 2B 3B 4C or 1C 2C 3C 4A. In contrast, the inheritance of either haplotype would be much less likely if the sites are far apart and could easily recombine by crossing over.

Haplotype Association Studies Are Conducted to Identify Disease-Causing Alleles

How do geneticists identify genes that cause disease when they are mutant? Although a variety of approaches may be followed, a common strategy is to localize a disease-causing allele to a small region on a chromosome that is distinguished by its haplotype. This approach, called a **haplotype association study,** is based on two assumptions:

1. The disease-causing allele had its origin in a single individual known as a **founder,** who lived many generations ago. Since that time, the allele has spread throughout portions of the human population.
2. When the disease-causing allele originated in the founder, it occurred in a region with a particular haplotype. The haplotype is not likely to have changed over the course of several generations if the alleles and markers in this region are very close together.

By comparing the transmission patterns of many molecular markers with the occurrence of an inherited disease, researchers can pinpoint particular markers that are closely linked to the disease-causing mutant gene.

To understand this concept, **Figure 23.6** shows a situation in which an individual—a founder—has incurred a new mutation that results in a disease-causing allele. The molecular markers designated 1A, 2C, 3B, and 4B are very close to the location of this mutant allele. After several generations, the disease-causing allele would more likely be present in individuals with the haplotype 1A 2C 3B 4B than in those without this haplotype.

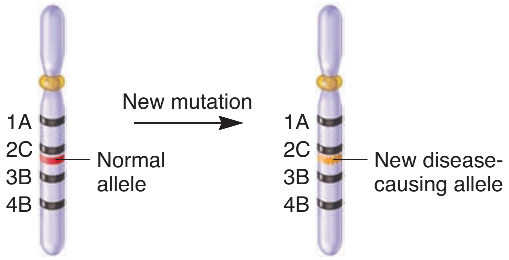


FIGURE 23.6 **The occurrence of a new mutation in a founder.** In this example, the new mutation occurred in a region with a specific haplotype: 1A 2C 3B 4B. Due to their close linkage, the same haplotype is likely to be found in individuals of succeeding generations who have inherited the mutant allele.

Furthermore, people who have inherited the disease-causing allele would be particularly likely to inherit the 2C marker because it is the closest to the disease-causing allele. Due to their close proximity, it would be unlikely that a crossover would separate the 2C marker and the disease-causing allele from each other and create a different haplotype. When alleles and molecular markers are associated with each other at a frequency that is significantly higher than expected by random chance, they are said to exhibit **linkage disequilibrium.** The phenomenon of linkage disequilibrium is common when a disease-causing allele arises in a founder and the allele is closely linked to other markers along a chromosome.

Haplotype Association Studies Identified the Mutant Gene That Causes Huntington Disease

As an example of a haplotype association study, let's consider Huntington disease. Nancy Wexler, whose mother died of this disorder, has studied Huntington disease among a population of related individuals in Venezuela. To date, over 15,000 individuals have been analyzed. **Figure 23.7** shows a simplified version of a large family pedigree from this Venezuelan population in which many individuals were affected with Huntington disease. A specific molecular marker, called G8, is located near the tip of the short arm of chromosome 4. It is found in four different versions, named A, B, C, or D. In this Venezuelan population, pedigree analysis revealed that the G8-C marker is almost always associated with the mutant gene causing Huntington disease. In other words, the G8-C marker is closely linked to the Huntington allele.

Once a disease-causing mutant gene has been localized to a short chromosomal region, the next step is to determine which gene in the region is responsible for the disease. Modern haplotype association studies typically localize a gene to a chromosome region that is about 1 Mb (1,000,000 bp) in length. One way to identify a disease-causing allele is chromosome walking (refer back to Figure 21.12). This is how the *Huntingtin* gene was identified. However, the technique of chromosome walking is no longer widely used because the entire human genome has been sequenced and most genes have been identified. Therefore, researchers can analyze the 1-Mb region where a gene has been mapped to determine if a mutant gene in this region is responsible for a disease. In the human genome, a 1-Mb region usually contains about 5 to 10 different genes, though the number can vary greatly. Therefore, mapping does not definitively tell researchers which gene may play a role in human disease, but it usually narrows down the list to a few candidate genes.

How does a researcher determine which of the candidate genes is the correct one? To further narrow down the list, researchers may also consider biological function. As the scientific community explores the functions of genes experimentally, the data are published in the research literature and placed into databases. In some cases, this information may help to narrow down the list of candidate genes. For example, if the disease of interest is

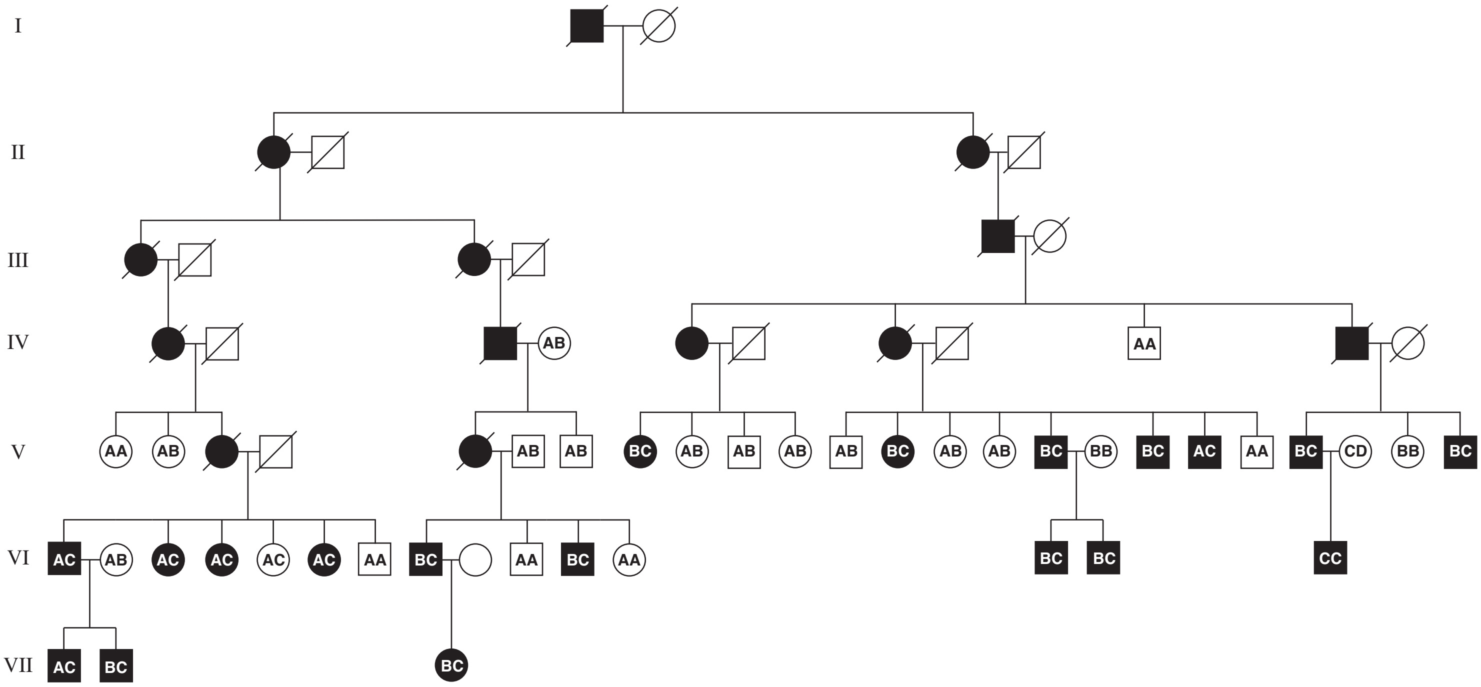


FIGURE 23.7 **The transmission pattern of a molecular marker for Huntington disease.** Affected members are shown with black symbols. The letters within each symbol indicate the forms of the G8 marker (A, B, C, or D) the individual carries. Affected individuals always carry the C version of this marker. Symbols with slashes indicate deceased individuals.

[Data from J. F. Gusella, N. S. Wexler, P. M. Conneally, et al. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306, 234–238.]

Concept check: *Explain the connection between the founder and the G8-C marker.*

neurological, researchers may discover that only certain genes in the mapped region are expressed in nerve cells. Also, researchers may compare data from other organisms. If a mutant gene in a mouse causes neurological problems and a human homolog of the mouse gene is found in the mapped region, this human gene would be a good candidate for being responsible for the disease symptoms.

After researchers have narrowed down the number of candidate genes to as short a list as possible, the next phase is to sequence the candidate gene(s) from many affected and unaffected individuals, using the DNA-sequencing methods described in Chapters 19 or 21. The goal is to identify a gene in which affected individuals always carry a mutation. Identifying a gene that has a genetic change found in affected individuals is strong evidence that the candidate gene is causing the disease symptoms.

Why is it useful to identify disease-causing mutant genes? In many cases, the identification of these genes helps us to understand how genes contribute to pathogenesis and may even aid in developing strategies aimed at the treatment of the disease. As described next, the identification of such genes may also result in the development of genetic tests that enable people to determine if they are carrying disease-causing alleles.

Genetic Testing Can Identify Many Inherited Human Diseases

Because genetic abnormalities occur in the human population at a significant level, people have sought ways to determine whether individuals possess disease-causing alleles. The term **genetic testing** refers to the use of testing methods to determine if an individual carries a particular allele, haplotype, or an alteration in chromosome structure or number. **Table 23.5** describes several different testing strategies. The term **genetic screening** refers to population-wide genetic testing.

In many cases, single-gene mutations that affect the function of cellular proteins can be examined at the protein level. If a gene encodes an enzyme, biochemical assays to measure that enzyme's activity may be available. As mentioned earlier, Tay-Sachs disease involves a defect in the enzyme hexosaminidase A (HexA). Enzymatic assays for this enzyme involve the use of an artificial substrate in which 4-methylumbelliferone (MU) is covalently linked to *N*-acetylglucosamine (GlcNAc). HexA cleaves this covalent bond and releases MU, which is fluorescent.

$$\underset{\text{(nonfluorescent)}}{\text{MU–GlcNAc}} \xrightarrow{\text{HexA}} \underset{\text{(fluorescent)}}{\text{MU}} + \text{GlcNAc}$$

To perform this assay, a small sample of cells is collected and incubated with MU–GlcNAc, and the fluorescence is measured with a device called a fluorometer. Individuals affected with Tay-Sachs, who do not produce the HexA enzyme, produce little or no fluorescence, whereas individuals who are homozygous for the normal *hexA* allele produce a high level of fluorescence. Heterozygotes, who have 50% HexA activity, produce intermediate levels of fluorescence.

An alternative and more common approach is to detect single-gene mutations at the DNA level. To apply this testing strategy, researchers must have previously identified the mutant gene using molecular techniques. The identification of many human genes, such as those involved in Duchenne muscular dystrophy, cystic fibrosis, and Huntington disease, has made it possible to test for affected individuals or those who may be carriers of these diseases. Table 23.5 describes several ways to test for gene mutations. The laboratory techniques used to test for these gene mutations are described in Chapters 19 and 21.

TABLE **23.5**

Testing Methods for Genetic Abnormalities

Method	Description
Protein Level	
Biochemical	As described in this chapter for Tay-Sachs disease, the enzymatic activity of a protein can be assayed in vitro.
Immunological	The presence of a protein can be detected using antibodies that specifically recognize that protein. Western blotting is an example of this type of technique (see Chapter 19).
DNA or Chromosomal Level	
RFLP analysis	RFLP analysis, which is described in Chapter 21, can be used to determine the likelihood that a person may carry a disease-causing allele.
DNA sequencing	If the normal gene has already been identified and sequenced, it is possible to design PCR primers that can amplify the gene from a sample of cells. The amplified DNA segment can then be subjected to DNA sequencing, as described in Chapter 19.
In situ hybridization	A DNA probe that hybridizes to a particular gene or gene segment can be used to determine if the gene is present, absent, or altered in an individual. The technique of fluorescence in situ hybridization (FISH) is described in Chapter 21 (see Figures 21.2, 21.3).
Karyotyping	The chromosomes from a sample of cells can be stained and then analyzed microscopically for abnormalities in chromosome structure and number (see Figure 2.2).
DNA microarrays	This method, which is described in Chapter 21, can be used to determine the expression levels of genes, such as those that are mutant in certain forms of cancer. In addition, microarrays are used to detect polymorphisms found in the human population that are associated with diseases.

Many human genetic abnormalities involve changes in chromosome number and/or structure. In fact, changes in chromosome number are a common class of human genetic abnormality. Most of these result in spontaneous abortions. However, approximately 1 in 200 live births have an abnormal number of chromosomes, a condition called aneuploidy (see Chapter 8, Table 8.1). About 5% of infant and childhood deaths are related to such genetic abnormalities. Changes in chromosome number and many changes in chromosome structure can be detected by karyotyping the chromosomes with a light microscope.

In the United States, genetic screening for certain disorders has become common medical practice. For example, pregnant women older than 35 are at higher risk of having children with chromosomal abnormalities, so they often have tests conducted to determine the chromosomal composition of their fetuses. Another example is the widespread screening of newborns for phenylketonuria (PKU). Those who test positive can then be given a low-phenylalanine diet to avoid PKU's devastating effects.

Genetic screening also has been conducted on specific populations in which a genetic disease is prevalent. For example, in 1971, community-based screening for heterozygous carriers of Tay-Sachs disease was begun among specific Ashkenazi Jewish populations. With the use of this screening, over the course of one generation, the incidence of TSD births was reduced by 90%. For most rare genetic abnormalities, however, genetic screening is not routine practice. Rather, genetic testing is performed only when a family history reveals a strong likelihood that a couple may produce an affected child. This typically involves a couple that already has an affected child or has other relatives with a genetic disease.

Genetic testing and screening are medical practices with many social and ethical dimensions. For example, people must decide whether or not they want to make use of available tests, particularly when the disease in question has no cure. For example, Huntington disease typically does not affect people until their 50s and can last 20 years. People who learn they are carriers of genetic diseases such as Huntington disease can be devastated by the news. Some argue that people have a right to know about their genetic makeup; others assert that knowing that one has an incurable, progressive disease does more harm than good. Another issue is privacy. Who should have access to personal genetic information, and how could it be used? Could routine genetic testing lead to discrimination by employers or medical insurance companies? In the coming years, we will gain an ever-increasing awareness of our genetic makeup and the underlying causes of genetic diseases. As a society, establishing guidelines for the uses of genetic testing will be a necessary, yet very difficult, task.

Genetic Testing Can Be Performed Prior to Birth

Testing for many genetic disorders can be performed during pregnancy. The two common ways of obtaining cellular material from a fetus for the purpose of genetic testing are amniocentesis and chorionic villus sampling. In **amniocentesis,** a doctor removes amniotic fluid containing fetal cells, using a needle that is passed through the abdominal wall (Figure 23.8). The fetal cells are cultured for several weeks and then subjected to genetic testing, such as karyotyping. In **chorionic villus sampling (CVS),** a small piece

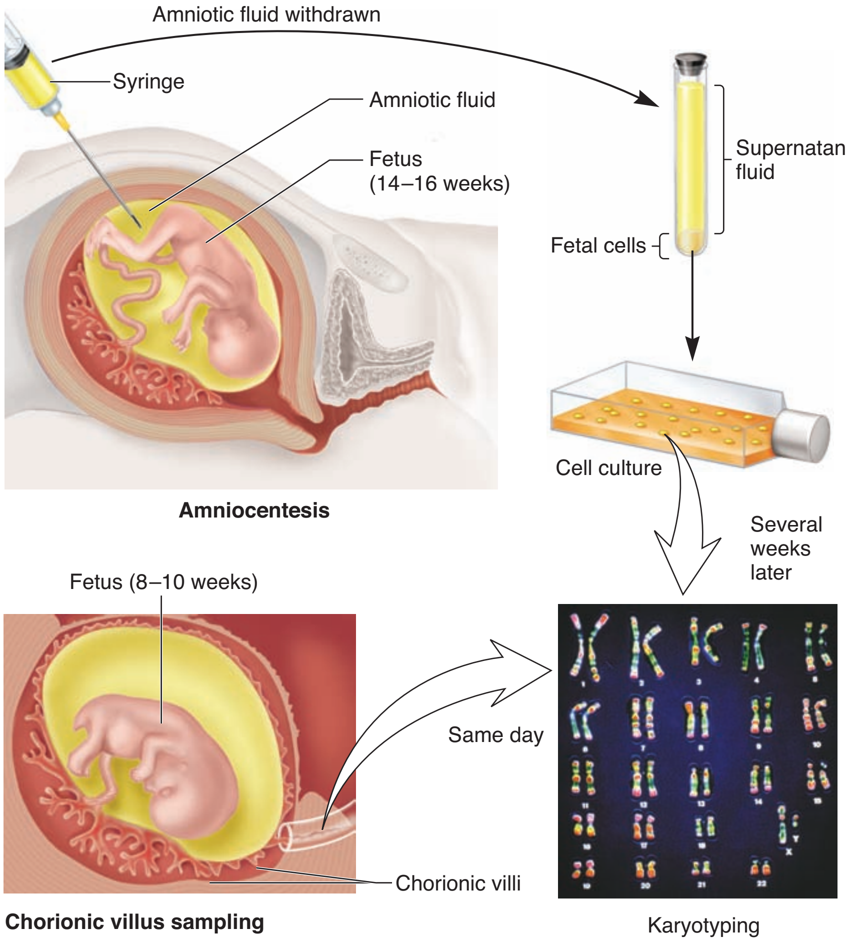


FIGURE 23.8 Techniques to determine genetic abnormalities during pregnancy. In amniocentesis, amniotic fluid is withdrawn, and fetal cells are collected by centrifugation. The cells are then allowed to grow in a laboratory culture medium for several weeks prior to karyotyping. In chorionic villus sampling, a small piece of the chorion is removed. These cells can be prepared directly for karyotyping.

of the chorion (the fetal part of the placenta) is removed, and genetic testing can be conducted directly on the collected cells. CVS can be performed earlier during pregnancy than amniocentesis, usually around the 8th to 10th week, compared with the 14th to 16th week for amniocentesis, and results are available sooner. Weighed against these advantages, however, is that CVS may pose a slightly greater risk of causing a miscarriage.

Another method of genetic screening prior to birth is called **preimplantation genetic diagnosis (PGD).** This approach, which is conducted before pregnancy even occurs, involves the genetic testing of embryos that have been produced by **in vitro fertilization (IVF)**—a process in which sperm and egg are combined outside of the mother's body. The testing is typically done to check for a specific genetic abnormality, such as the allele that causes Huntington disease. PGD can also determine if an embryo contains the correct number of chromosomes (also called aneuploidy screening).

PGD is done by removing one or two cells usually at about the eight-cell stage, which occurs 3 days after fertilization. This process is called embryo biopsy or blastomere biopsy. Molecular techniques described in Table 23.5 are then conducted on the removed cell(s) to either check for a particular genetic disease or determine the chromosome composition. The testing can usually be completed in a day or so. Depending on the outcome of the results, a decision can be made whether or not to transfer the embryo into the uterus of the prospective mother in hopes of implantation and the eventual birth of a baby. In most cases, only embryos that do not harbor genetic abnormalities are used. As with the genetic screening of adults, the screening of embryos and fetuses raises many ethical questions.

23.2 REVIEWING THE KEY CONCEPTS

- The term *haplotype* refers to the linkage of alleles or molecular markers along a single chromosome (see Figure 23.5).
- Disease-causing mutations may occur in a founder with a specific haplotype (see Figure 23.6).
- Researchers may identify disease-causing alleles by determining their location within a chromosome via haplotype association studies (see Figure 23.7).
- Genetic testing for abnormalities can be performed in a variety of ways (see Table 23.5).
- Genetic screening is population-wide genetic testing.
- Genetic testing can be done prior to birth, which includes amniocentesis, chorionic villus sampling, and preimplantation genetic diagnosis (see Figure 23.8).

23.2 COMPREHENSION QUESTIONS

1. What is a haplotype?
   a. A species with one set of chromosomes
   b. A cell with one set of chromosomes
   c. The linkage of alleles or molecular markers along a chromosome
   d. All of the above
2. Haplotype association studies are aimed at the identification of a particular _________ based on ______________.
   a. chromosome, an abnormality in its structure
   b. chromosome, the arrangement of molecular markers
   c. gene, its linkage to other genes or molecular markers
   d. gene, chromosomal rearrangements
3. Which of the following is *not* a method used in genetic testing?
   a. Chromosome walking
   b. DNA sequencing
   c. In situ hybridization
   d. DNA microarrays
4. Which of the following prenatal genetic testing methods is done in conjunction with in vitro fertilization?
   a. Amniocentesis
   b. Chorionic villus sampling
   c. Preimplantation genetic diagnosis
   d. All of the above

23.3 PRIONS

We now turn to an unusual mechanism in which agents known as prions cause disease. Such diseases can be inherited or they may be acquired by eating infected meat. As shown in **Table 23.6**, prions cause several types of neurodegenerative diseases affecting humans and livestock, including bovine spongiform encephalopathy, or mad cow disease. Recent evidence has shown that prions also exist in yeast. In the 1960s, British researchers Tikvah Alper and John Stanley Griffith discovered that preparations from animals with certain neurodegenerative diseases remained infectious even after exposure to radiation that would destroy any DNA or RNA. They suggested that the infectious agent was a protein. Furthermore, Alper and Griffith speculated that the protein usually preferred one folding pattern but could sometimes misfold and then catalyze other proteins to do the same. In the early 1970s, Stanley Prusiner, moved by the death of a patient from a neurodegenerative disease, began to search for the causative agent. In 1982, he isolated a disease-causing agent composed entirely of protein, which he called a **prion.** The term emphasizes the prion's unusual character as a proteinaceous infectious agent. Before the discovery of prions, all known infectious agents such as viruses and bacteria contained their own genetic material (either DNA or RNA).

Prion-related diseases arise from the ability of the prion protein to exist in two conformational states: a normal form, PrP^{C}, which does not cause disease, and an abnormal form, PrP^{Sc}, which does. (Note: The superscript C refers to the normal conformation, and the superscript Sc refers to the abnormal conformation, such as the one found in the disease called scrapie.) The gene encoding the prion protein (*PrP*) is found in humans and other mammals, and the protein is expressed at low levels in certain types of cells such as nerve cells. The abnormal conformation of the prion protein can come from two sources. An individual can be infected with the abnormal protein by taking the abnormal protein into their bodies. For example, someone may eat products from an animal that had the disease. Alternatively, some people carry alleles of the *PrP* gene that cause their prion protein to convert spontaneously to the abnormal conformation at a very low rate. These individuals have an inherited

TABLE **23.6**

Neurodegenerative Diseases Caused by Prions*

Disease	Description
Infectious Diseases	
Kuru	A human disease that was once common in New Guinea. It begins with a loss of coordination, usually followed by dementia. Infection was spread by cannibalism, a practice that ended in 1958.
Scrapie	A disease of sheep and pigs characterized by intense itching in which the animals tend to scrape themselves against trees, followed by neurodegeneration
Mad cow disease	Begins with changes in posture and temperament, followed by loss of coordination and neurodegeneration
Human Inherited Diseases	
Creutzfeldt-Jakob disease	Characterized by loss of coordination and dementia
Gerstmann-Straüssler-Scheinker disease	Characterized by loss of coordination and dementia
Familial fatal insomnia	Begins with sleeping and autonomic nervous system disturbances followed by insomnia and dementia

*All of these diseases are eventually fatal.

predisposition to develop a prion-related disease. An example of an inherited prion disease is familial fatal insomnia (Table 23.6).

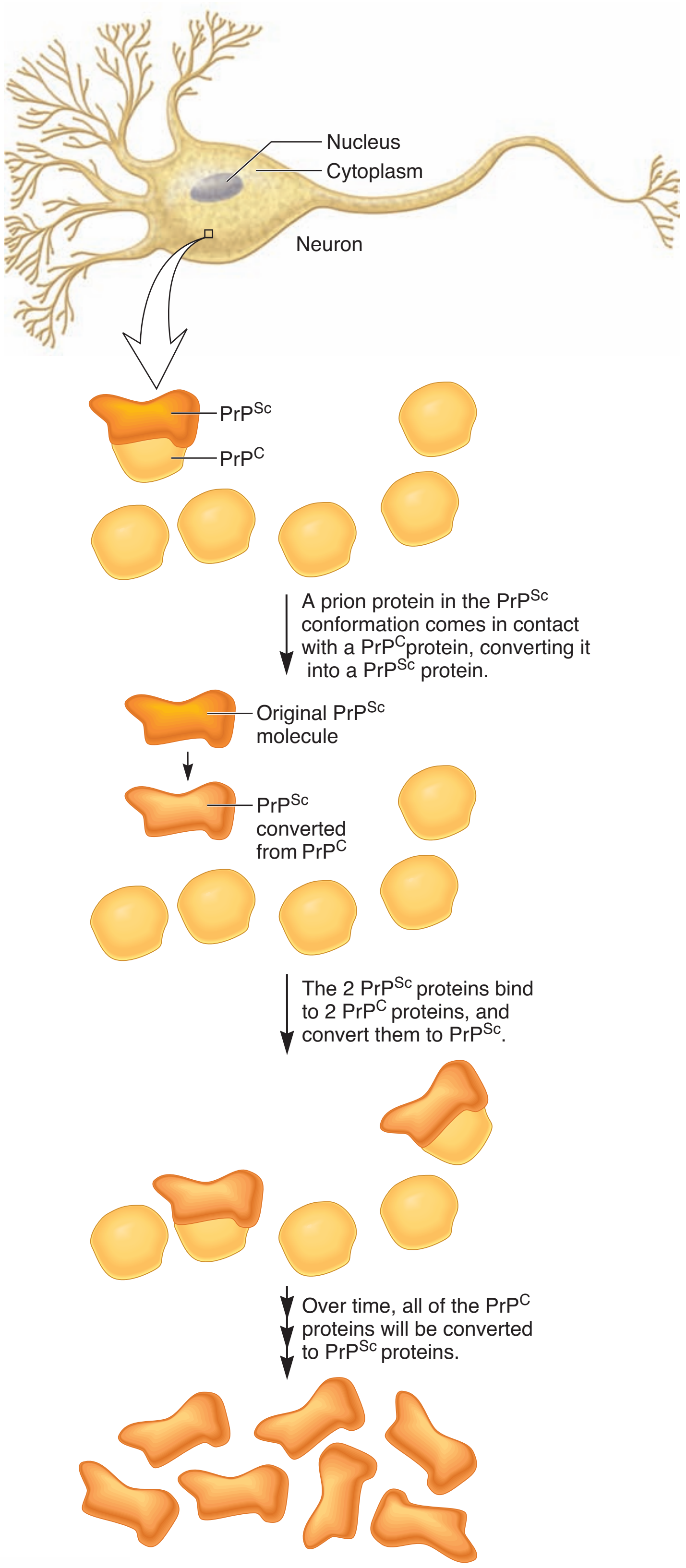
What is the molecular mechanism through which prions cause disease? As noted, the prion protein can exist in two conformations, PrP^{C} and PrP^{Sc}. As shown in **Figure 23.9**, the abnormal conformation, PrP^{Sc}, acts as a catalyst to convert normal prion proteins within the cell to the misfolded conformation. As a prion disease progresses, the PrP^{Sc} proteins form dense aggregates in the cells of the brain and peripheral nervous tissues. This deposition is correlated with the disease symptoms affecting the nervous system. Some of the abnormal prion proteins are also excreted from infected cells and travel through the nervous system to infect other cells.

23.3 REVIEWING THE KEY CONCEPTS

- Prions are infectious agents composed entirely of protein. They can be acquired by eating infected meat or via a genetic predisposition (see Table 23.6).
- Prions cause disease because the PrP protein in the abnormal (prion) conformation is able to convert normal proteins to the abnormal conformation (see Figure 23.9).

23.3 COMPREHENSION QUESTIONS

1. A prion is a disease-causing agent composed of
   a. cells.
   b. nucleic acid and protein coat.
   c. protein alone.
   d. nucleic acid alone.


Nucleus
Cytoplasm
Neuron
PrP^{Sc}
PrP^{C}
A prion protein in the PrP^{Sc} conformation comes in contact with a PrP^{C} protein, converting it into a PrP^{Sc} protein.
Original PrP^{Sc} molecule
PrP^{Sc} converted from PrP^{C}
The 2 PrP^{Sc} proteins bind to 2 PrP^{C} proteins, and convert them to PrP^{Sc}.
Over time, all of the PrP^{C} proteins will be converted to PrP^{Sc} proteins.

ONLINE ANIMATION

FIGURE 23.9 A proposed molecular mechanism of prion diseases. The PrP^{Sc} protein catalyzes the conversion of PrP^{C} to PrP^{Sc}. Over time, the PrP^{Sc} conformation accumulates to high levels, leading to symptoms of the prion disease.

Concept check: Where does the PrP^{Sc} protein come from?

23.4 GENETIC BASIS OF CANCER

Cancer is a disease characterized by uncontrolled cell division. More than 100 kinds of human cancers have been identified, and they are classified according to the type of cell that has become cancerous. Though cancer is a diverse collection of many diseases, some characteristics are common to all cancers.

1. Most cancers originate in a single cell. This single cell, and its line of daughter cells, usually undergoes a series of genetic changes that accumulate during cell division. In this regard, a cancerous growth can be considered to be **clonal** in origin. A hallmark of a cancer cell is that it divides to produce two daughter cancer cells.
2. At the cellular and genetic levels, cancer usually is a multistep process that begins with a precancerous genetic change—a **benign** growth—and is followed by additional genetic changes that lead to cancerous cell growth (Figure 23.10).
3. When cells have become cancerous, their growth is described as **malignant.** Malignant cells have three common characteristics: (1) cell division continues in an unlimited manner; (2) the cells are **invasive,** which means they can invade healthy tissues; and (3) the cells are **metastatic,** which means they can migrate to other parts of the body and cause secondary tumors.

In the United States, approximately 1 million people are diagnosed with cancer each year, and about half that number will die from the disease. In 5% to 10% of all cases, a predisposition to develop the cancer is an inherited trait. We will examine some inherited forms of cancer later in this chapter. Most cancers, though, perhaps 90% to 95%, are not passed from parent to offspring. Rather, cancer is usually an acquired condition that typically occurs later in life. Although some cancers are caused by spontaneous mutations and viruses, at least 80% of all human cancers are related to exposure to agents that promote genetic changes in somatic cells. These environmental agents, such as UV light and certain chemicals, are mutagens that alter the DNA in a way that affects the function of normal genes. If the DNA is permanently modified in somatic cells, such changes may be transmitted during cell division. These DNA alterations can lead to effects on gene expression that ultimately affect cell division and thereby lead to cancer. An environmental agent that causes cancer in this manner is called a **carcinogen.**

In this section, we begin by considering the general features of cell division in eukaryotes. We will then explore how genetic abnormalities, which affect the functions of particular cellular proteins, can lead to cancer.

The Eukaryotic Cell Cycle Is Controlled by Growth Factors Pathways

To understand how mutations cause cancer, we first need to examine the normal process of cell division. As described in Chapter 2, eukaryotic cells destined to divide progress through a series of stages known as the cell cycle (Figure 23.11a). The phases consist of G_1 (first gap), S (synthesis of DNA, the genetic material), G_2 (second gap), and M phase (mitosis and cytokinesis). The G_1 phase is a period in a cell's life when it may become committed to divide. Depending on the conditions, a cell in the G_1 phase may accumulate molecular changes that cause it to progress through the rest of the cell cycle. When this occurs, cell biologists say that a cell has reached a special control point called the restriction point—a point in the cell cycle when a cell becomes committed to progress through the rest of the cell cycle. The commitment to divide is based on a variety of factors.

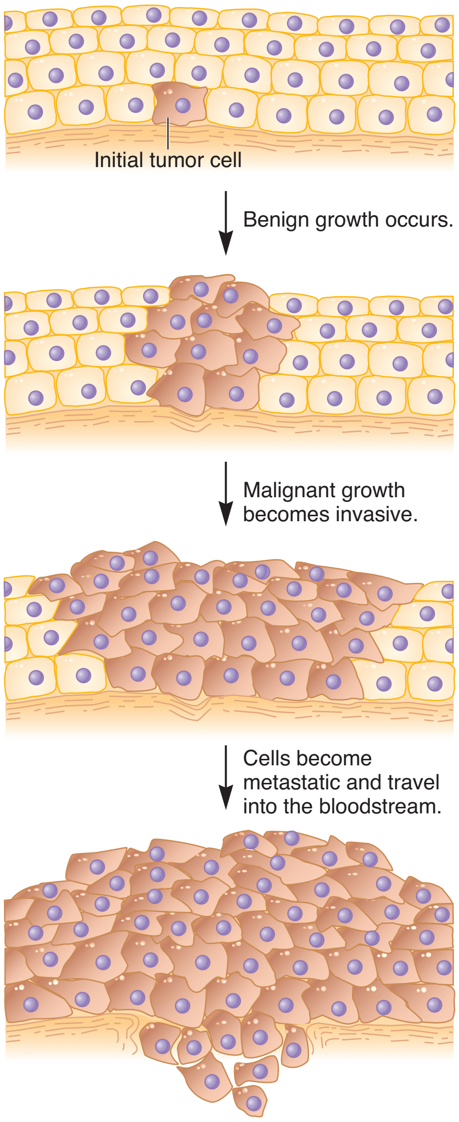

Initial tumor cell
Benign growth occurs.
Malignant growth becomes invasive.
Cells become metastatic and travel into the bloodstream.

FIGURE 23.10 **Progression of cellular growth leading to cancer.**

Genes→Traits In a healthy individual, an initial gene mutation converts a normal cell into a tumor cell. This tumor cell divides many times to produce a benign tumor. Additional genetic changes in the tumor cells may occur, eventually leading to malignant growth. At a later stage in malignancy, the tumor cells invade surrounding tissues, and some malignant cells may metastasize by traveling through the bloodstream to other parts of the body, where they can grow and cause secondary tumors. As a trait, cancer can be viewed as a series of genetic changes that eventually lead to uncontrolled cell growth.

For example, environmental conditions, such as the presence of sufficient nutrients, are important for cell division. In addition, multicellular organisms rely on signaling molecules to coordinate cell division throughout the body. These signaling molecules are often called **growth factors** because they promote cell division.

As researchers began to study cancer at the cellular level, they wanted to understand how these mutant genes promote abnormal cell division. In parallel with cancer research, cell biologists have studied the roles that normal cellular proteins play in cell division. As mentioned, the cell cycle is regulated in part by growth factors that bind to cell surface receptors and initiate a cascade of cellular events that lead eventually to cell division. **Figure 23.11b** considers a protein called epidermal growth factor (EGF) that is secreted from endocrine cells and stimulates epidermal cells, such as skin cells, to divide. As seen here, EGF binds to its receptor, leading to the activation of an intracellular signaling pathway. This pathway, also known as a signal cascade, leads to a change in gene transcription. In other words, the transcription of specific genes is activated in response to the growth hormone. Once these genes are transcribed and translated, the gene products function to promote the progression through the cell cycle. Figure 23.11b is just one example of a pathway between a growth factor and gene activation. Eukaryotic species produce many different growth factors, and the signaling pathways are often more complex than the one shown here.

Oncogenes Have Gain-of-Function Mutations That May Affect Proteins Involved in Cell Division Pathways

What is the relationship between normal genes and mutant genes that cause cancer? Let's first consider a category of cancer-causing genes called oncogenes. An **oncogene** is an abnormally active gene that promotes cancerous growth. A normal, nonmutated gene that has the potential to become an oncogene is termed a **proto-oncogene.** To become an oncogene, a proto-oncogene must incur a mutation that causes its expression to be abnormally active. This is another example of a gain-of-function mutation, which we discussed earlier in this chapter in the context of dominant mutations. A gain-of-function mutation that produces an oncogene typically has one of three possible effects:

1. The amount of the encoded protein is greatly increased.
2. A change occurs in the structure of the encoded protein that causes it to be overly active.
3. The encoded protein is expressed in a cell type where it is not normally expressed.

The mutations that convert proto-oncogenes into oncogenes have been analyzed in many types of cancers. Oncogenes commonly encode proteins that function in cell growth signaling pathways (**Table 23.7**). These include growth factors, growth factor receptors, proteins involved in the intracellular signaling pathways, and transcription factors. Let's consider two examples. Research groups headed by Robert Gallo and Mark Groudine showed that a *myc* gene, designated c-*myc*, was amplified about 10-fold in a human promyelocytic leukemia cell line. Since that time, research has shown that *myc* genes are overexpressed in many forms of cancer, including breast cancer, lung cancer, and colon carcinoma. The overexpression of this transcription factor leads to the transcriptional activation of genes that promote cell division.

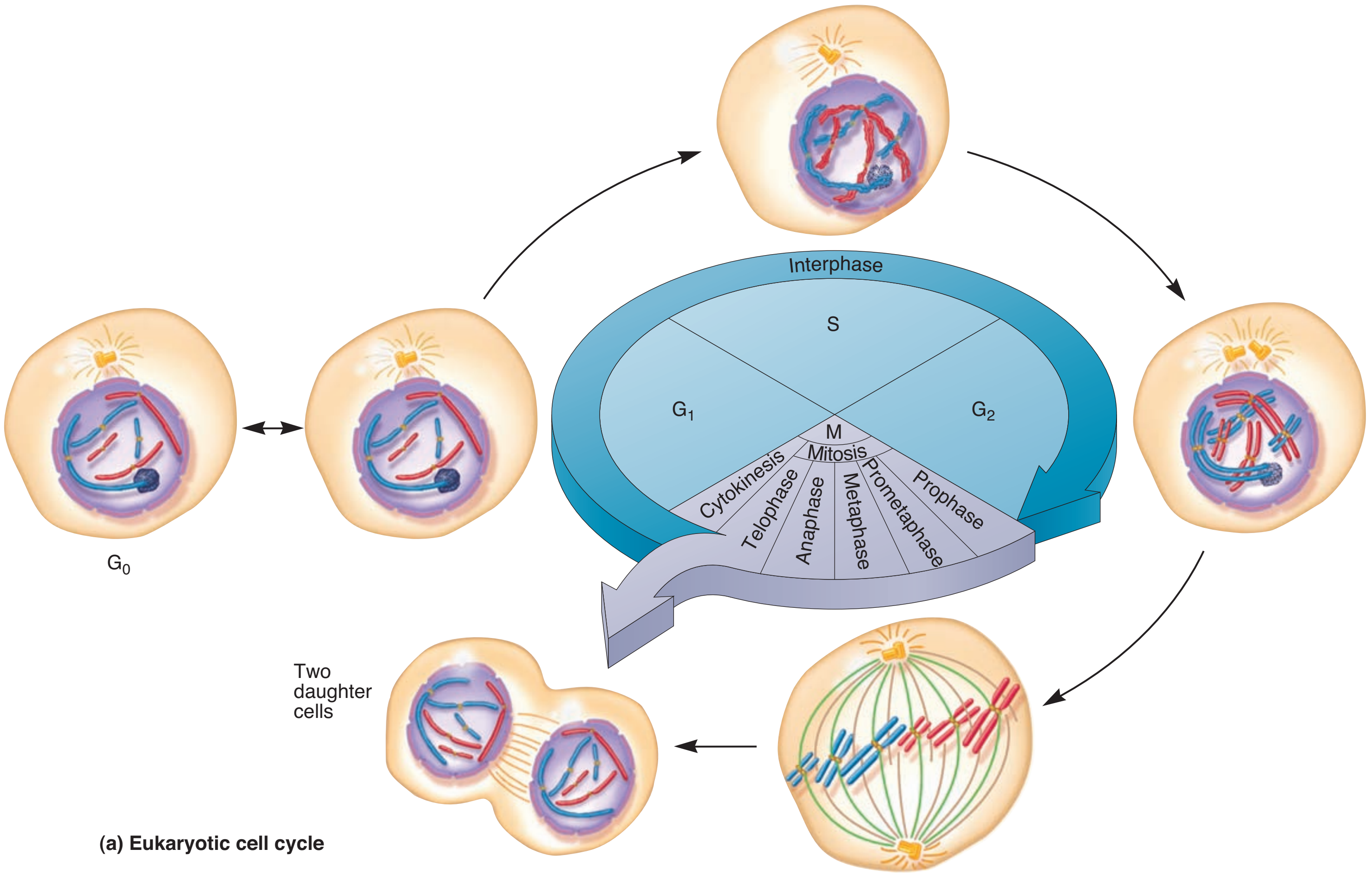
As a second example, mutations that alter the amino acid sequence of the Ras protein have been shown to cause functional abnormalities. The Ras protein is a GTPase, which hydrolyzes GTP to GDP + P_i (**Figure 23.12**). Therefore, after it has been activated, the Ras protein returns to its inactive state by hydrolyzing GTP. Mutations that convert the normal *ras* gene into an oncogene either decrease the ability of the Ras protein to hydrolyze GTP or increase the rate of exchange of bound GDP for GTP. Both of these functional changes result in a greater amount of the active GTP-bound form of the Ras protein. In this way, these mutations keep the signaling pathway turned on, thereby stimulating the cell to divide.

TABLE 23.7

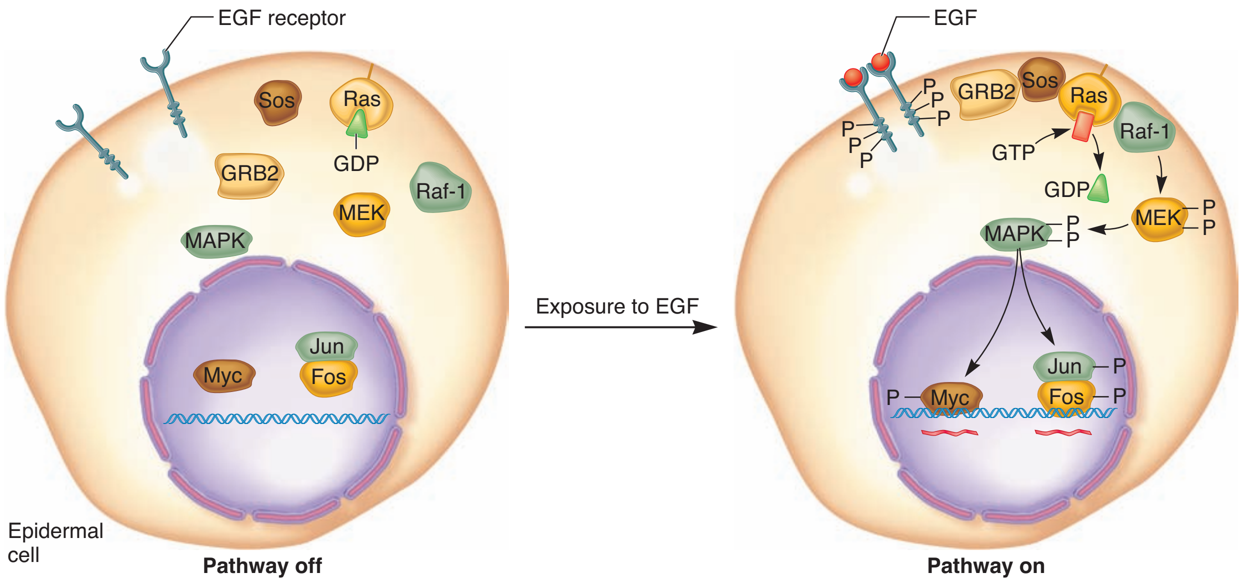
Examples of Proto-Oncogenes That Can Mutate into Oncogenes

Gene	Cellular Function
Growth Factors*	
sis	Platelet-derived growth factor
int-2	Fibroblast growth factor
Growth Factor Receptors	
erbB	Growth factor receptor for EGF (epidermal growth factor)
fms	Growth factor receptor for NGF (nerve growth factor)
K-sam	Growth factor receptor for FGF (fibroblast growth factor)
Intracellular Signaling Proteins	
ras	GTP/GDP-binding protein
raf	Serine/threonine kinase
src	Tyrosine kinase
abl	Tyrosine kinase
gsp	G-protein α subunit
Transcription Factors	
myc	Transcription factor
jun	Transcription factor
fos	Transcription factor
gli	Transcription factor
erbA	Transcription factor (steroid receptor)

*The genes described in this table are found in humans as well as other vertebrate species. Many of these genes were initially identified in retroviruses. Most of the genes have been given three-letter names that are abbreviations for the type of cancer the oncogene causes or the type of virus in which the gene was first identified.


Interphase
S
G_1
G_2
M
Mitosis
Cytokinesis
Telophase
Anaphase
Metaphase
Prometaphase
Prophase
G_0
Two daughter cells

(a) Eukaryotic cell cycle


EGF receptor
Sos
Ras
GDP
GRB2
Raf-1
MEK
MAPK
Jun
Myc
Fos
Epidermal cell
Pathway off
Exposure to EGF
EGF
P
GRB2
Sos
Ras
Raf-1
GTP
GDP
MAPK
MEK
Jun
Myc
Fos
Pathway on

(b) EGF signaling pathway

FIGURE 23.11 **The eukaryotic cell cycle and activation of a cell-signaling pathway by a growth factor. (a)** The cell cycle involves a progression through the G_1, S, G_2, and M phases. Progression through the cell cycle is often stimulated by growth factors. (Note: Chromosomes are not condensed during G_0, G_1, S, and G_2 phases. They are shown that way in this figure so they can be counted.) **(b)** In this example, epidermal growth factor (EGF) binds to two EGF receptors, causing them to dimerize and phosphorylate each other. An intracellular protein called GRB2 is attracted to the phosphorylated EGF receptor, and it is subsequently bound by another protein called Sos. The binding of Sos to GRB2 enables Sos to activate a protein called Ras. This activation involves the binding of GTP and the release of GDP. The activated Ras/GTP complex then activates Raf-1, which is a protein kinase. Raf-1 phosphorylates MEK, and then MEK phosphorylates MAPK. More than one MAPK may be involved. Finally, the phosphorylated form of MAPK activates transcription factors, such as Myc, Jun, and Fos. This leads to the transcription of genes, which encode proteins that promote cell division.

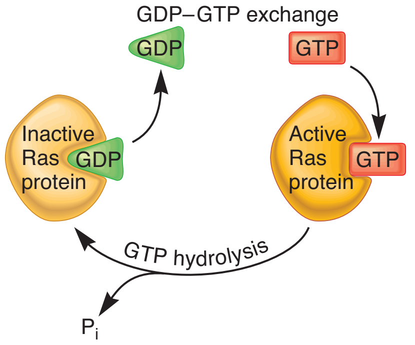


FIGURE 23.12 **Functional cycle of the Ras protein.** The binding of GTP to Ras activates the function of Ras and promotes cell division. The hydrolysis of GTP to GDP and P_i converts the active form of Ras to an inactive form.

Concept check: *How would a mutation that prevents the Ras protein from hydrolyzing GTP affect the EFG pathway described in Figure 23.11b?*

TABLE 23.8

Examples of Genetic Changes That Convert Proto-Oncogenes into Oncogenes

Type of Change	Description and Examples
Missense mutation	A change in the amino acid sequence of a proto-oncogene protein may cause it to function in an abnormal way. Missense mutations can convert *ras* genes into oncogenes.
Gene amplification	The copy number of a proto-oncogene may be increased by gene duplication. *Myc* genes have been amplified in human leukemias; breast, stomach, lung, and colon carcinomas; and neuroblastomas and glioblastomas. *ErbB* genes have been amplified in glioblastomas, squamous cell carcinomas, and breast, salivary gland, and ovarian carcinomas.
Chromosomal translocation	A piece of chromosome may be translocated to another chromosome and affect the expression of genes at the breakpoint site. In Burkitt lymphoma, a region of chromosome 8 is translocated to either chromosome 2, 14, or 22. The breakpoint in chromosome 8 causes the overexpression of the c-*myc* gene.
Viral integration	When a virus integrates into a chromosome, it may enhance the expression of nearby proto-oncogenes. In avian lymphomas, the integration of the avian leukosis virus can enhance the transcription of the c-*myc* gene.

Genetic Changes in Proto-Oncogenes Convert Them to Oncogenes

How do specific genetic alterations convert proto-oncogenes into oncogenes? By isolating and studying oncogenes at the molecular level, researchers have discovered four main ways this occurs (**Table 23.8**). These changes can be categorized as missense mutations, gene amplifications (i.e., an increase in copy number), chromosomal translocations, and viral integrations, which are described next.

Missense Mutations As mentioned previously, changes in the structure of the Ras protein can cause it to become permanently activated. These changes are caused by a missense mutation in the *ras* gene. The human genome contains four different but evolutionarily related *ras* genes: *ras*H, *ras*N, *ras*K-4a, and *ras*K-4b. All four homologous genes encode proteins with very similar amino acid sequences containing a total of 188 or 189 amino acids. Missense mutants in these normal *ras* genes are associated with particular forms of cancer. For example, a missense mutation in *ras*H that changes a glycine to a valine is responsible for the conversion of *ras*H into an oncogene:

	1	2	3	4	5	6	7	8	9	10	11	12	13		188	189
Normal	Met	Thr	Glu	Tyr	Lys	Leu	Val	Val	Val	Gly	Ala	Gly	Gly		Leu	Ser
Human *ras*H	ATG	ACG	GAA	TAT	AAG	CTG	GTG	GTG	GTG	GGC	GCC	GGC	GGT	....	CTC	TCC
												↓				
Oncogenic												GTC				
*ras*H	Met	Thr	Glu	Tyr	Lys	Leu	Val	Val	Val	Gly	Ala	Val	Gly	....	Leu	Ser

Experimentally, chemical carcinogens, which are discussed in Chapter 18, have been shown to cause this type of missense mutation, thereby leading to cancer. This is a gain-of-function mutation that causes the Ras protein to be abnormally active, thereby promoting cell division.

Gene Amplification Another genetic event that may occur in cancer cells is gene amplification, or an abnormal increase in the copy number of a proto-oncogene. An increase in gene copy number is expected to increase the amount of the encoded protein, thereby contributing to malignancy. Gene amplification does not normally happen in mammalian cells, but it is a common occurrence in cancer cells. As mentioned previously, Gallo and Groudine discovered that c-*myc* was amplified in a human leukemia cell line. Many human cancers are associated with the amplification of particular oncogenes. In such cases, the extent of oncogene amplification may be correlated with the progression of tumors to increasing malignancy. These include the amplification of N-*myc* in neuroblastomas and *erbB-2* in breast carcinomas. In other types of malignancies, gene amplification is more random and may be a secondary event that increases the expression of oncogenes previously activated by other genetic changes.

Chromosomal Translocation A third type of genetic alteration that can lead to cancer is a chromosomal translocation. Although structural abnormalities are common in cancer cells, very specific types of chromosomal translocations have been identified in certain types of tumors. In 1960, Peter Nowell and David Hungerford discovered that chronic myelogenous leukemia (CML) is correlated with the presence of a shortened version of chromosome 22, which they called the Philadelphia chromosome after the city where it was discovered. This shortened chromosome is the result of a reciprocal translocation between chromosomes 9 and 22. Later studies revealed that this translocation activates a proto-oncogene, *abl*, in an unusual way (**Figure 23.13**). The reciprocal translocation involves breakpoints within the *abl* and *bcr* genes. Following the reciprocal translocation, the coding sequence of the *abl* gene fuses with the promoter and coding sequence of the *bcr* gene. This yields an oncogene that encodes an abnormal fusion protein, which contains the polypeptide sequences encoded from both genes. The *abl* gene encodes a tyrosine kinase enzyme, which uses ATP to attach phosphate groups onto target proteins. This phosphorylation activates certain proteins involved with cell division. Normally, the *abl* gene is highly regulated. However, in the Philadelphia chromosome, the *bcr* promoter, which is active in white blood cells, controls the expression of the fusion gene. This leads to an overexpression of the tyrosine kinase function in such cells. This explains why this fusion causes a type of cancer called a leukemia, which involves a proliferation of white blood cells.

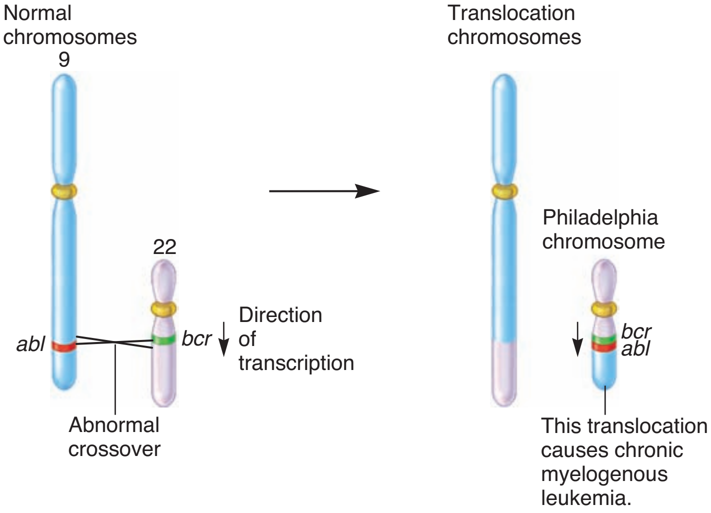

Normal chromosomes
9
22
abl
bcr
Direction of transcription
Abnormal crossover
Translocation chromosomes
Philadelphia chromosome
bcr
abl
This translocation causes chronic myelogenous leukemia.

FIGURE 23.13 The reciprocal translocation commonly found in people with chronic myelogenous leukemia.

Genes→Traits In healthy individuals, the *abl* gene is located on chromosome 9, and the *bcr* gene is on chromosome 22. In chronic myelogenous leukemia, a reciprocal translocation causes the *abl* gene to fuse with the *bcr* gene. This combined gene, under the control of the *bcr* promoter, encodes an abnormal fusion protein that overexpresses the tyrosine kinase function of the ABL protein and leads to leukemia.

Concept check: Why does this translocation cause leukemia rather than cancer in a different tissue type, such as the lung?

Interestingly, the study of the *abl* gene has led to an effective treatment for CML. Until recently, the only successful treatment was to destroy the patient's bone marrow and then restore blood-cell production by infusing stem cells from the bone marrow of a healthy donor. With knowledge about the function of the ABL protein, researchers have developed the drug imatinib mesylate (Gleevec) that appears to dramatically improve survival. This molecule fits into the active site of the ABL protein, preventing ATP from binding there. Without ATP, the ABL protein cannot phosphorylate its target proteins. This prevents the ABL protein from stimulating cell division. In a clinical trial, almost 90% of the CML patients treated with the drug showed no further progression of their disease!

Other forms of cancer also involve chromosomal translocations that cause an overexpression of an oncogene. In Burkitt lymphoma, for example, a region of chromosome 8 is translocated to chromosome 2, 14, or 22. The breakpoint in chromosome 8 is near the c-*myc* gene, and the sites on chromosomes 2, 14, and 22 correspond to locations of different immunoglobulin genes that are normally expressed in lymphocytes. The translocation of the c-*myc* gene near the immunoglobulin genes leads to the overexpression of the c-*myc* gene, thereby promoting malignancy in lymphocytes.

Viral Integration A fourth way that oncogenes can occur is via viral integration. As part of their reproductive cycle, certain viruses, such as retroviruses, integrate their genomes into the chromosomal DNA of their host cell. If the integration occurs next to a proto-oncogene, a viral promoter or enhancer sequence may cause the proto-oncogene to be overexpressed. For example, in certain lymphomas that occur in birds, the genome of the avian leukosis virus was found to be integrated next to the c-*myc* gene and enhanced its level of transcription.

Certain Viruses Cause Cancer by Carrying Viral Oncogenes into the Cell

As mentioned, most types of cancers are caused by mutagens that alter the structure and expression of genes. A few viruses, however, are known to carry oncogenes that cause cancer in plants, animals, and humans. Most cancer-causing viruses are not very potent at inducing cancer. An organism must be infected for a long time before a tumor actually develops. By comparison, a few types of viruses rapidly induce tumors in animals and efficiently transform cells in culture. These are called **acutely transforming viruses (ACTs).** About 40 such viruses have been isolated from chickens, turkeys, mice, rats, cats, and monkeys. The first virus of this type, the Rous sarcoma virus (RSV), was isolated from chicken sarcomas by Peyton Rous in 1911.

During the 1970s, RSV research led to the identification of the first gene known to cause cancer. Researchers investigated RSV by infecting chicken fibroblast cells in the laboratory. This causes the chicken fibroblasts to grow like cancer cells. During the course of their studies, researchers identified mutant RSV strains that infected and proliferated within chicken cells without transforming them into malignant cells. These RSV strains were determined to contain a defective viral gene. In contrast, in other strains where this gene is functional, cancer occurs. This viral gene was designated *src* for sarcoma, the type of cancer it causes. The *src* gene is known as v-*src* because it is found within a viral genome.

Because the viral oncogene of RSV is not necessary for viral replication, researchers were curious as to why the virus has this v-*src* oncogene. Harold Varmus and Michael Bishop, in collaboration with Peter Vogt, soon discovered that normal host cells contain a copy of the *src* gene in their chromosomes. This normal copy of the *src* gene found in the chromosomal DNA of the host is termed c-*src*, for cellular *src*. The c-*src* gene does not cause cancer. However, once incorporated into a viral genome, this gene can become a viral oncogene that promotes cancer. Why does cancer occur when the gene is in a viral genome? Three explanations of this phenomenon are possible. First, the many copies of the virus made during viral replication may lead to overexpression of the *src* gene. Second, the incorporation of the *src* gene next to viral regulatory sequences may cause it to be overexpressed. A third possibility is that the v-*src* gene may accumulate additional mutations that convert it to an oncogene.

RSV has acquired the *src* gene by capturing it from a host cell's chromosome. This can occur during the RSV life cycle. RSV is a retrovirus with an RNA genome. During its life cycle, a retrovirus uses reverse transcriptase to make a DNA copy of its genome, which becomes integrated as a provirus into the host cell genome. This integration could occur next to the c-*src* gene. During transcription of the proviral DNA, the *src* gene could be included in the RNA transcript that becomes the retroviral genome.

Since these early studies of RSV, many other retroviruses carrying oncogenes have been investigated. The characterization of such oncogenes has led to the identification of several genes with cancer-causing potential. In addition to retroviruses, several viruses with DNA genomes cause tumors, and some of these are known to cause cancer in humans (**Table 23.9**). Researchers estimate that up to 15% of all human cancers are associated with viruses.

TABLE **23.9**

Examples of Viruses That May Cause Cancer

Virus	Description
Retroviruses	
Rous sarcoma virus	Causes sarcomas in chickens.
Simian sarcoma virus	Causes sarcomas in monkeys.
Hardy-Zuckerman-4 feline sarcoma virus	Causes sarcomas in cats.
DNA Tumor Viruses	
Hepatitis B, polyomavirus	Causes liver cancer in several species including humans. In some cases, these viruses do not cause cancer in their natural hosts but can transform cells in culture.
Human papillomavirus (HPV)	Causes benign tumors and malignant carcinomas in several species, including humans. Causes cervical cancer in humans.
Herpesvirus	Causes carcinoma in frogs and T-cell lymphoma in chickens. A human herpesvirus, Epstein-Barr virus, is a causative agent in Burkitt lymphoma, which occurs primarily in immunosuppressed individuals such as AIDS patients.

Tumor-Suppressor Genes Play a Role in Preventing the Proliferation of Cancer Cells

As we have seen, an oncogene is an abnormally activated gene that leads to uncontrolled cell growth. We now turn our attention to a second category of genes called **tumor-suppressor genes.** As the name suggests, the role of a tumor-suppressor gene is to prevent cancerous growth. Therefore, when a tumor-suppressor gene becomes inactivated by mutation, it becomes more likely that cancer will occur. It is a loss-of-function mutation in a tumor-suppressor gene that promotes cancer.

The first identification of a human tumor-suppressor gene involved studies of retinoblastoma, a tumor that occurs in the retina of the eye. Some people have inherited a predisposition to develop this disease within the first few years of life. By comparison, the noninherited form of retinoblastoma, which is caused by environmental agents, tends to occur later in life but only rarely.

Based on these differences, in 1971, Alfred Knudson proposed a "two-hit" model for retinoblastoma. According to this model, retinoblastoma requires two mutations to occur. People with the hereditary form already have received one mutant gene from one of their parents. They need only one additional mutation in the other copy of this tumor-suppressor gene to develop the disease. Because the retina has more than 1 million cells, it is relatively likely that a mutation may occur in one of these cells at an early age, leading to the disease. However, people with the noninherited form of the disease must have two mutations in the same retinal cell to cause the disease. Because two rare events are much less likely to occur than a single such event, the noninherited form of this disease is expected to occur much later in life and only rarely. Therefore, this hypothesis explains the different populations typically affected by the inherited and noninherited forms of retinoblastoma.

Since Knudson's original hypothesis, molecular studies have confirmed the two-hit hypothesis for retinoblastoma. In this case, the gene in which mutations occur is designated *rb* (for retinoblastoma). This tumor suppressor gene is found on the long arm of chromosome 13. Most people have two normal copies of the *rb* gene. Persons with hereditary retinoblastoma have inherited one normal and one defective copy. In nontumorous cells throughout the body, they have one functional copy and one defective copy of *rb*. However, in retinal tumor cells, the normal *rb* gene has also suffered the second hit (i.e., a mutation), which renders it defective. Without the tumor-suppressor ability provided by the Rb protein, cells are allowed to grow and divide in an unregulated manner, which ultimately leads to cancer. (In contrast, as discussed later, most other forms of cancer involve mutations in several genes.)

More recent studies have revealed how the Rb protein suppresses the proliferation of cancer cells (**Figure 23.14**). The Rb protein regulates a transcription factor called E2F, which activates genes required for cell cycle progression. (The eukaryotic cell cycle is described earlier in Figure 23.11a.) The binding of the Rb protein to E2F inhibits its activity and prevents the cell from progressing through the cell cycle. As discussed later in this chapter, when a normal cell is supposed to divide, cellular proteins called cyclins bind to cyclin-dependent protein kinases (CDKs). This activates the CDKs, which then leads to the phosphorylation of the Rb protein. The phosphorylated form of the Rb protein is released from E2F, thereby allowing E2F to activate genes needed to progress through the cell cycle. What happens when both copies of the *rb* gene are rendered inactive by mutation? The answer is that the E2F protein is always active, which explains why uncontrolled cell division occurs.

The Vertebrate *p53* Gene Is a Master Tumor-Suppressor Gene That Senses DNA Damage

After the *rb* gene, the second tumor-suppressor gene discovered was the *p53* gene. The *p53* gene is the most commonly altered gene in human cancers. About 50% of all human cancers are associated with defects in *p53*. These include malignant tumors of the lung, breast, esophagus, liver, bladder, and brain as well as sarcomas, lymphomas, and leukemias. For this reason, an enormous amount of research has been aimed at elucidating the function of the p53 protein.

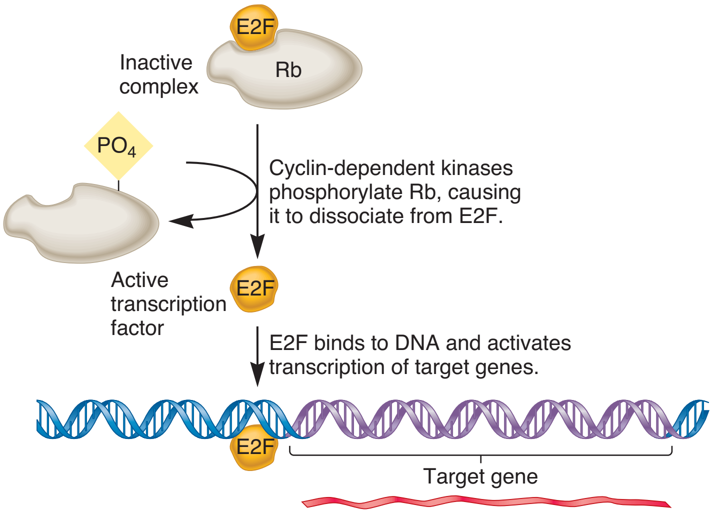

E2F
Inactive complex
Rb
PO4
Cyclin-dependent kinases phosphorylate Rb, causing it to dissociate from E2F.
Active transcription factor
E2F
E2F binds to DNA and activates transcription of target genes.
E2F
Target gene

FIGURE 23.14 **Interactions between the Rb and E2F proteins.** The binding of the Rb protein to the transcription factor E2F inhibits the ability of E2F to function. This prevents cell division. For cell division to occur, cyclins bind to cyclin-dependent protein kinases, which then phosphorylate the Rb protein. The phosphorylated Rb protein is released from E2F. The free form of E2F can activate target genes needed to progress through the cell cycle.

Concept check: *If a cell cannot make any Rb protein, how will this affect the function of E2F?*

A primary role of the p53 protein is to determine if a cell has incurred DNA damage. If damage is detected, p53 can promote three types of cellular pathways aimed at preventing the proliferation of cells with damaged DNA. First, when confronted with DNA damage, the cell can try to repair its DNA. This may prevent the accumulation of mutations that activate oncogenes or inactivate tumor-suppressor genes. Second, if a cell is in the process of dividing, it can arrest itself in the cell cycle. By stopping the cell cycle, a cell has more time to repair its DNA and avoid producing two mutant daughter cells. For this to happen, p53 stimulates the expression of another gene termed *p21*. The p21 protein inhibits cyclin/CDK protein complexes that are needed to progress from the G_1 phase of the cell cycle to the S phase.

The third, and most drastic event, is that a cell can initiate a series of events called **apoptosis,** or programmed cell death. In response to DNA-damaging agents, a cell may self-destruct. Apoptosis is an active process that involves cell shrinkage, chromatin condensation, and DNA degradation. This process is facilitated by proteases known as **caspases.** These types of proteases are sometimes called the “executioners” of the cell. Caspases digest selected cellular proteins such as microfilaments, which are components of the intracellular cytoskeleton. This causes the cell to break down into small vesicles that are eventually phagocytized by cells of the immune system. Apoptosis occurs in some cells as a normal process of embryonic development (see Chapter 24). In addition, it is an important way by which an adult organism can eliminate cells with cancer-causing potential.

Figure 23.15 summarizes how *p53* plays a central role in all three processes. As shown here, the expression of the *p53* gene is caused by the formation of damaged DNA. The inducing signal appears to be double-strand DNA breaks. The p53 protein functions as a transcription factor. It contains a DNA-binding domain and a transcriptional activation domain. Experimental studies have shown that p53 can activate the transcription of several specific target genes. As shown in Figure 23.15, it can activate genes that promote DNA repair, arrest the cell cycle, and promote apoptosis. In addition, p53 appears to act as a negative regulator by interacting with general transcription factors, which decreases the general expression of many other structural genes. This inhibition may also help prevent the cell from dividing.

Overall, p53 activates the expression of a few specific cellular genes and, at the same time, inhibits the expression of many other genes in the cell. Through the regulation of gene transcription, the p53 protein can prevent the proliferation of cells that have incurred DNA damage.

Tumor-Suppressor Genes Encode Proteins that Negatively Regulate Cell Division or Maintain Genome Integrity

During the past few decades, researchers have identified many tumor-suppressor genes that, when defective, contribute to the

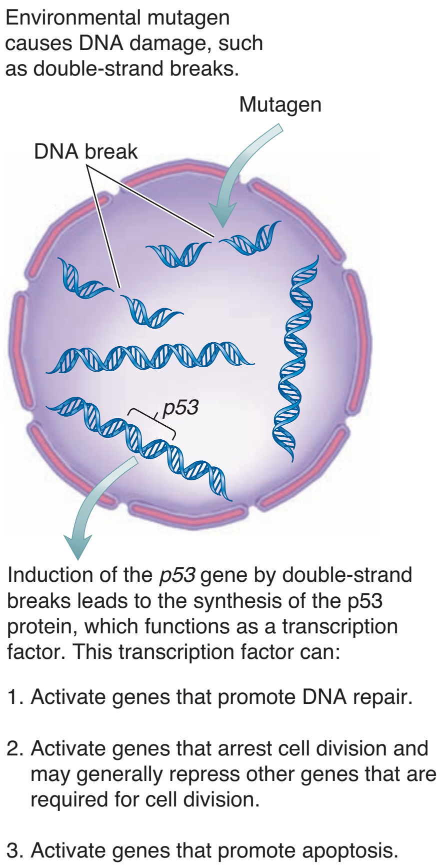

Environmental mutagen causes DNA damage, such as double-strand breaks.
Mutagen
DNA break
p53
Induction of the *p53* gene by double-strand breaks leads to the synthesis of the p53 protein, which functions as a transcription factor. This transcription factor can:
1. Activate genes that promote DNA repair.
2. Activate genes that arrest cell division and may generally repress other genes that are required for cell division.
3. Activate genes that promote apoptosis.

FIGURE 23.15 **Central role of *p53* in preventing the proliferation of cancer cells.** The *p53* gene, which encodes a transcription factor, is induced by cellular proteins that sense double-strand DNA breaks. The p53 transcription factor may activate genes that promote DNA repair, activate genes that arrest cell division and repress genes required for cell division, and activate genes that promote apoptosis.

TABLE 23.10

Functions of Selected Tumor-Suppressor Genes

Gene	Function
Genes That Negatively Regulate Cell Division	
rb	The Rb protein is a negative regulator of E2F (see Figure 23.14). The inhibition of E2F prevents the transcription of certain genes required for DNA replication and cell division.
p16	A protein kinase that negatively regulates cyclin-dependent kinases. This protein controls the transition from the G_1 phase of the cell cycle to the S phase.
NF1	The NF1 protein stimulates Ras to hydrolyze its GTP to GDP. Loss of NF1 function causes the Ras protein to be overactive, which promotes cell division.
APC	A negative regulator of a cell-signaling pathway called the Wnt pathway. The Wnt pathway leads to the activation of genes that promote cell division.
VHL	VHL protein functions in the targeting of certain proteins for (ubiquitin-mediated) protein degradation. When these proteins are not degraded, the cell cycle may proceed in an unregulated manner.
MTS1	The MTS1 protein acts as an inhibitor of cyclin-dependent protein kinases. The activities of these protein kinases promote the progression through the cell cycle.
Genes That Maintain Genome Integrity	
p53	p53 is a transcription factor that is a checkpoint protein that positively regulates a few specific target genes and negatively regulates others in a general manner. It acts as a sensor of DNA damage. It can prevent the progression through the cell cycle and also can promote apoptosis.
BRCA-1, *BRCA-2*	BRCA1 and BRCA2 proteins are both involved in the cellular defense against DNA damage. These proteins facilitate DNA repair and act as checkpoint proteins.

development and progression of cancer (**Table 23.10**). What are the general functions of these genes? They tend to fall into two broad categories: genes that negatively regulate cell division or genes that maintain genome integrity.

Some tumor-suppressor genes encode proteins that have direct effects on the regulation of cell division. The *rb* gene is an example. As mentioned earlier, the Rb protein negatively regulates E2F. If both copies of the *rb* gene are inactivated, the growth of cells is accelerated. Therefore, loss of function of these kinds of negative regulators has a direct effect on the abnormal cell division rates seen in cancer cells. In other words, when the Rb protein is lost, a cell becomes more likely to divide.

Alternatively, other tumor-suppressor genes play a role in the proper maintenance of the integrity of the genome. The term **genome maintenance** refers to cellular mechanisms that either prevent mutations from occurring and/or prevent mutant cells from surviving or dividing. The proteins encoded by such genes help to ensure that gene mutations or changes in chromosome structure and number do not occur and are not transmitted to daughter cells. The proteins that participate in genome maintenance can be subdivided into two classes: checkpoint proteins and those involved directly with DNA repair.

The role of many **checkpoint proteins** is to detect genetic abnormalities, such as DNA breaks and improperly segregated chromosomes. When such abnormalities are detected, the proteins participate in regulatory pathways that prevent cell division. These proteins are called checkpoint proteins because their role is to check the integrity of the genome and prevent cells from progressing past a certain point in the cell cycle if genetic abnormalities are detected.

Figure 23.16 shows a simplified diagram of cell cycle control. Proteins called cyclins and cyclin-dependent protein kinases (CDKs) are responsible for advancing a cell through the four phases of the cell cycle. For example, an activated G_1-cyclin/CDK complex is necessary to advance from the G_1 phase to the S phase. Human cells produce several types of cyclins and CDKs. The formation of activated cyclin/CDK complexes is regulated by a variety of factors. Earlier we considered how growth factor signaling pathways can stimulate cell division (see Figure 23.11b). These pathways operate, in part, by regulating cyclins and CDKs. In addition, checkpoint proteins monitor the state of the cell and stop the progression through the cell cycle if abnormalities are

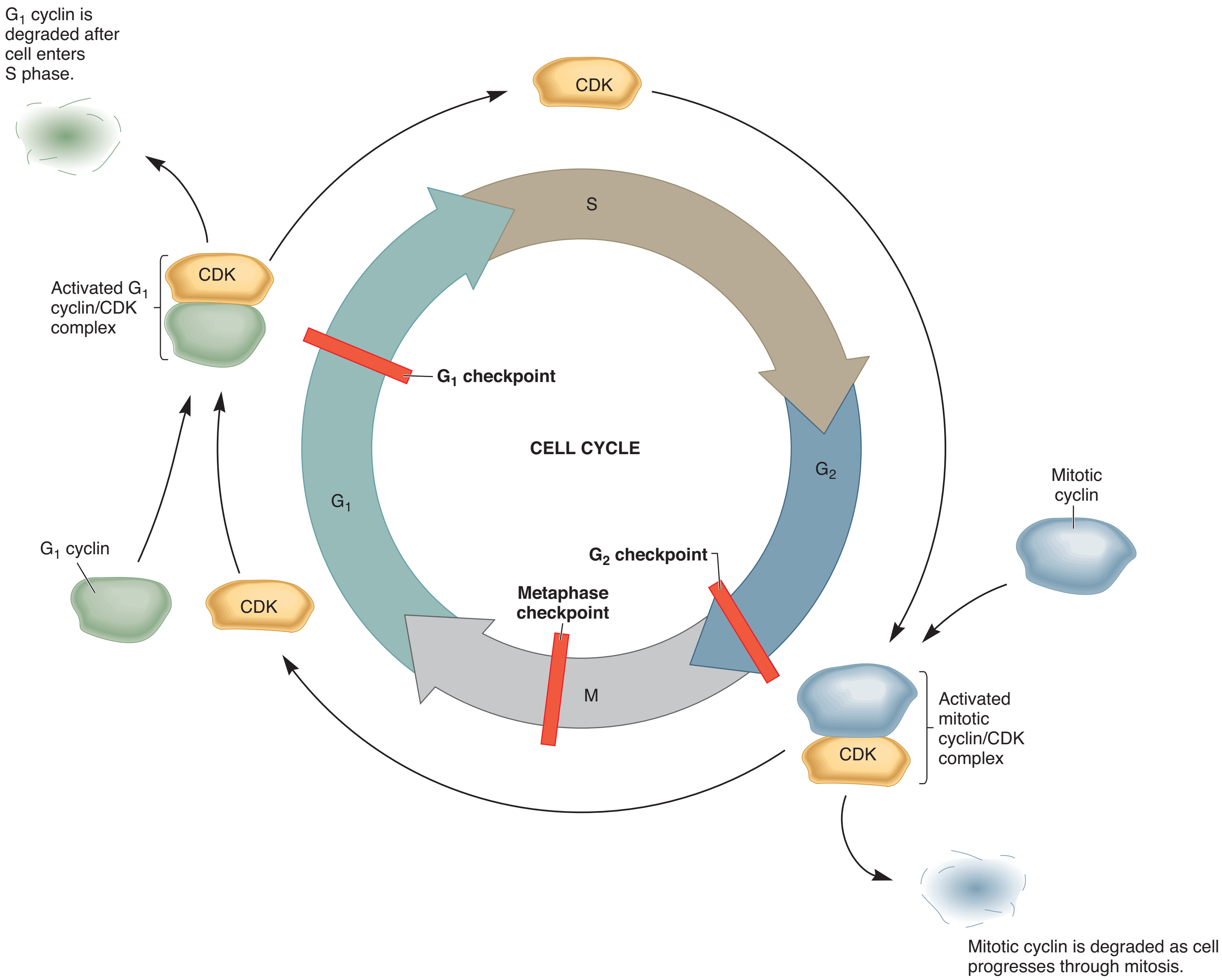

G$_1$ cyclin is degraded after cell enters S phase.
CDK
S
Activated G$_1$ cyclin/CDK complex
CDK
G$_1$ checkpoint
CELL CYCLE
G$_2$
Mitotic cyclin
G$_1$
G$_1$ cyclin
CDK
G$_2$ checkpoint
Metaphase checkpoint
M
Activated mitotic cyclin/CDK complex
CDK
Mitotic cyclin is degraded as cell progresses through mitosis.

FIGURE 23.16 **Cell cycle control.** Progression through the cell cycle requires the formation of activated cyclin/CDK complexes. A cell may produce several different types of cyclin proteins, which are typically degraded after the cell has progressed to the next phase. The formation of activated cyclin/CDK complexes is regulated by checkpoint proteins. If these proteins detect DNA damage, they will prevent the formation of activated cyclin/CDK complexes. In addition, other checkpoint proteins can detect chromosomes that are incorrectly attached to the spindle apparatus and prevent further progression through metaphase. Note that this is a simplified diagram of the cell cycle of humans.

Concept check: *What is a checkpoint?*

detected. Checkpoint proteins also may exert their effects by regulating the activities of cyclins and CDKs.

Several checkpoint proteins regulate the cell cycle of human cells. Figure 23.16 shows three of the major checkpoints where these proteins exert their effects. Both the G$_1$ and G$_2$ checkpoints involve the functions of proteins that can sense if the DNA has incurred damage. If so, these checkpoint proteins, such as p53, can prevent the formation of active cyclin/CDK complexes. This stops the progression of the cell cycle. A checkpoint also exists in metaphase. This checkpoint is monitored by proteins that can sense if a chromosome is not correctly attached to the spindle apparatus, making it likely that it will be improperly segregated.

Overall, checkpoint proteins prevent the division of cells that may have incurred DNA damage or harbor abnormalities in chromosome attachment. This provides a mechanism to stop the accumulation of genetic abnormalities that could produce cancer cells within the body. When checkpoint genes are lost, cell division may not be directly accelerated. However, the loss of checkpoint protein function makes it more likely that undesirable genetic changes occur that could cause cancerous growth.

A second class of proteins involved with genome maintenance consists of DNA repair enzymes, which were discussed in Chapter 18 (refer back to Table 18.7). The genes encoding such enzymes are inactivated in certain forms of cancer. The loss of a DNA repair enzyme makes it more likely for a cell to accumulate mutations that could create an oncogene or eliminate the function of a tumor-suppressor gene. In Chapter 18, we considered how defects in the nucleotide excision repair process are responsible for the disease called xeroderma pigmentosum, which results in a predisposition to developing skin cancer (refer back to Figure 18.20). As discussed later in this chapter, defects in DNA mismatch repair enzymes can contribute to colorectal cancer. In these cases, the loss of a DNA repair enzyme contributes to a higher mutation rate, which makes it more likely for other genes to incur cancer-causing mutations.

Tumor-Suppressor Genes Can Be Silenced in a Variety of Ways

Thus far, we have considered the functions of proteins that are encoded by tumor-suppressor genes. Cancer biologists also want to understand how tumor-suppressor genes are inactivated, because this knowledge may aid in the prevention of cancer. Researchers have identified three common ways that the function of tumor-suppressor genes can be lost. First, a mutation can occur specifically within a tumor-suppressor gene to inactivate its function. For example, a mutation could inactivate the promoter of a tumor-suppressor gene or introduce an early stop codon in the coding sequence. Either of these would prevent the expression of a functional protein. A second way that tumor-suppressor genes are inhibited is via DNA methylation. As discussed in Chapter 17, DNA methylation usually inhibits the transcription of eukaryotic genes, particularly when it occurs in the vicinity of the promoter. The methylation of CpG islands near the promoters of tumor-suppressor genes has been found in many types of tumors, suggesting that this form of gene inactivation plays an important role in the formation or progression of malignancy. However, further research is needed to determine why tumor-suppressor genes are aberrantly methylated in cancer cells. Third, many types of cancer are associated with aneuploidy. As discussed in Chapter 8, aneuploidy involves the loss or addition of one or more chromosomes, so the total number of chromosomes is not an even multiple of a set. In some cases, chromosome loss may contribute to the progression of cancer because the lost chromosome carries one or more tumor-suppressor genes.

Most Forms of Cancer Involve Multiple Genetic Changes Leading to Malignancy

The discovery of oncogenes and tumor-suppressor genes, along with molecular techniques that can detect genetic alterations, has enabled researchers to study the progression of certain forms of cancer at the molecular level. Many cancers begin with a benign genetic alteration that, over time and with additional mutations, eventually progresses to malignancy. Furthermore, a malignancy can continue to accumulate genetic changes that make it even more difficult to treat. For example, some tumors may acquire mutations that cause them to be resistant to chemotherapeutic agents.

In 1990, Eric Fearon and Bert Vogelstein proposed a series of genetic changes that lead to colorectal cancer, the second most common cancer in the United States. As shown in **Figure 23.17**, colorectal cancer is derived from cells in the mucosa of the colon. The loss of function of *APC*, a tumor-suppressor gene on chromosome 5, leads to an increased proliferation of mucosal cells and the development of a benign polyp, a noncancerous growth. Additional genetic changes involving the loss of other tumor-suppressor genes and the activation of an oncogene (namely, *ras*) lead to the development of a carcinoma. In Figure 23.17, the genetic changes that cause colon cancer are portrayed as occurring in an orderly sequence. Although the growth of a tumor often begins with mutations in *APC*, it is the total number of genetic changes, not their exact order, that is important. A key issue is that multiple mutations are needed to disable checkpoints.

Among different types of tumors, researchers have identified a large number of genes that are mutated in cancer cells. Though not all of these mutant genes have been directly shown to affect the growth rate of cells, such mutations are likely to be found in tumors because they provide some type of growth advantage for the cell population from which the cancer developed. For example, certain mutations may enable cells to metastasize to neighboring locations. These mutations may not affect growth rate, but they provide the growth advantage that cancer cells are not limited to growing in a particular location, but can migrate to new locations.

Researchers have estimated that about 300 different genes may play a role in the development of human cancer. With an approximate genome size of 20,000 to 25,000 genes, this observation indicates that over 1% of our genes have the potential to promote cancer if their function is altered by a mutation.

In addition to mutations within specific genes, other common genetic changes associated with cancer are abnormalities in chromosome structure and number. **Figure 23.18** compares the chromosome composition of a normal male cell and a tumor cell taken from the same person. The normal composition for this person is 22 pairs of chromosomes plus two sex chromosomes (X and Y). By comparison, the chromosome composition of the tumor cell is quite bizarre, including the observation that the tumor cell has two X chromosomes, which is characteristic of females. There are many cases where chromosomes are missing. If tumor-suppressor genes were on these missing chromosomes, their function is lost as well. Figure 23.18 also shows a few cases of extra chromosomes. If these chromosomes contain proto-oncogenes, the expression of those genes may be overactive. Finally, tumor cells often contain chromosomes that have translocations. Such translocations may create fused genes (as in the case of the Philadelphia chromosome discussed earlier in this chapter), or they may place two genes close together so that the regulatory sequences of one gene affect the expression of the other.

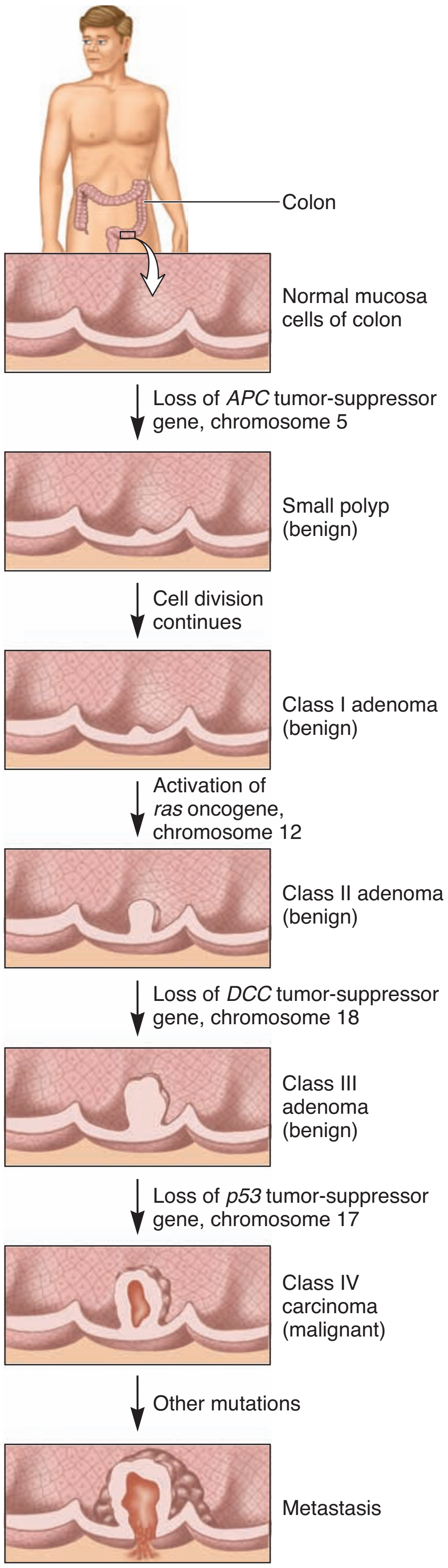

Colon
Normal mucosa cells of colon
Loss of *APC* tumor-suppressor gene, chromosome 5
Small polyp (benign)
Cell division continues
Class I adenoma (benign)
Activation of *ras* oncogene, chromosome 12
Class II adenoma (benign)
Loss of *DCC* tumor-suppressor gene, chromosome 18
Class III adenoma (benign)
Loss of *p53* tumor-suppressor gene, chromosome 17
Class IV carcinoma (malignant)
Other mutations
Metastasis

FIGURE 23.17 **Multiple genetic changes leading to colorectal cancer.**

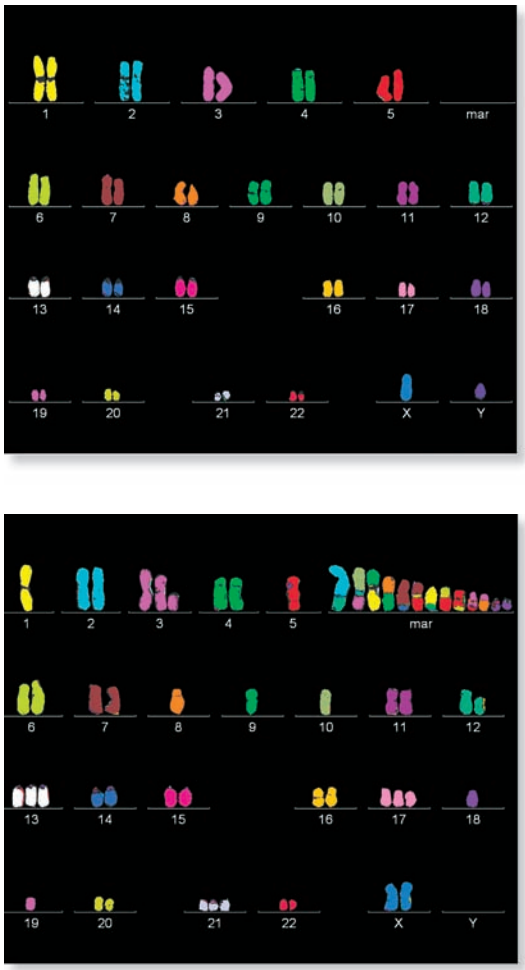

1
2
3
4
5
mar
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

FIGURE 23.18 **A comparison between chromosomes found in a normal human cell and a cancer cell from the same person.** The bottom set found in a cancer cell is highly abnormal, with extra copies of some chromosomes and lost copies of others. Chromosomes made of fused pieces of chromosomes (designated mar in this figure) are also common in cancer cells.

Molecular Profiling Is Increasingly Used to Classify Tumors

As we have just seen, cancer cells are usually the result of multiple genetic alterations that cause the activation of oncogenes and the loss of function of tumor-suppressor genes. Therefore, each type of tumor is characterized by a particular set of gene and chromosome alterations. Traditionally, different types of tumors have not been identified on the basis of genetic changes but instead have been classified according to their appearance under a microscope. Although this approach is useful, a major drawback is that two tumors may have a very similar microscopic appearance but yet have very different underlying genetic changes and clinical outcomes. For this reason, researchers and clinicians are turning to methods that enable them to understand the molecular changes that occur in diseases such as cancer. This general approach is called **molecular profiling.** For example, as described in Chapter 21, DNA microarrays are used to study the expression of many genes simultaneously. Cancer cells usually carry multiple genetic changes that can affect the expression of many genes, not just those that have become mutant. DNA microarrays are used to classify the pattern of gene expression that occurs in different types of tumors.

Why is molecular profiling useful? First, it can distinguish between tumors that look very similar under the microscope. Second, as we gain a better understanding of the molecular changes associated with particular types of cancers, researchers may be able to develop drugs that specifically target the proteins that are encoded by cancer-causing gene mutations or drugs that target nonmutant proteins whose expression is altered in the malignant state. As discussed earlier, the drug imatinib mesylate, which is used to treat chronic myelogenous leukemia, was developed in this way.

A third benefit of molecular profiling is that it may affect treatment options and patient outcome. For example, about 70% of all breast cancers exhibit an overexpression of the estrogen receptor. These types of breast cancer are better treated with drugs that either block the estrogen receptor or block the synthesis of estrogen. An example is tamoxifen, which is an antagonist of the estrogen receptor. The application of genetic or molecular data in the treatment of disease is called **personalized medicine.** As we gain a better understanding of human genes and disease states, researchers expect that personalized medicine will become an increasingly important aspect of health care.

Inherited Forms of Cancers May Be Caused by Defects in Tumor-Suppressor Genes

Before we end our discussion of the genetic basis of cancer, let's consider which genes are most likely to be affected in inherited forms of the disease. As mentioned previously, about 5% to 10% of all cases of cancer involve inherited (germ-line) mutations. These familial forms of cancer occur because people have inherited mutations from one or both parents that give them an increased susceptibility to developing cancer. This does not mean they will definitely get cancer, but they are more likely to develop the disease than are individuals in the general population. When individuals have family members who have developed certain forms of cancer, they may be tested to determine if they also carry a mutant gene. For example, von Hippel-Lindau disease and familial adenomatous polyposis are examples of syndromes for which genetic testing to identify at-risk family members is considered the standard of care.

What types of genes are mutant in familial cancers? Most inherited forms of cancer involve a defect in a tumor-suppressor gene (**Table 23.11**). In these cases, the individual is heterozygous, with one normal and one inactive allele.

At the phenotypic level, a predisposition for developing cancer is inherited in a dominant fashion because a heterozygote exhibits this predisposition. **Figure 23.19a** shows a pedigree for familial breast cancer. In this case, individuals with the disorder have inherited a loss-of-function mutation in the *BRCA-1* gene. As seen in the pedigree, the development of breast cancer shows a dominant pattern of inheritance with incomplete penetrance, which means that heterozygotes do not always exhibit the trait. Most affected individuals have an affected parent. However, the actual development of cancer is recessive, because it initially relies on the loss of function of the normal copy of the *BRCA-1* gene from a somatic cell (**Figure 23.19b**). This phenomenon is called **loss of heterozygosity (LOH)**—the loss of function of a normal allele when the other allele was already inactivated. As discussed earlier, LOH also explains the inherited form of retinoblastoma.

Loss of heterozygosity occurs in a variety of ways. For example, the normal *BRCA-1* gene could suffer a point mutation or a deletion that inactivates its function. Alternatively, the

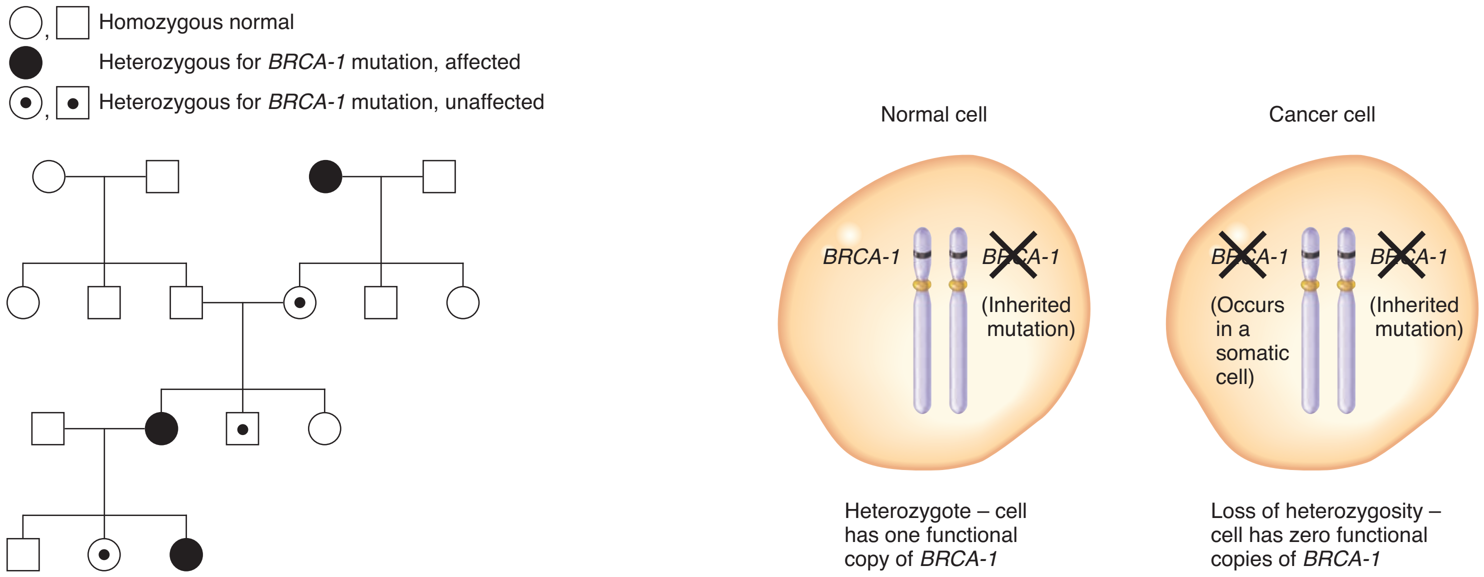
TABLE 23.11

Inherited Mutant Genes That Confer a Predisposition to Develop Cancer

Gene	Type of Cancer*
Tumor-Suppressor Genes†	
VHL	Causes von Hippel-Lindau disease, which is typically characterized by a clear-cell renal carcinoma
APC	Familial adenomatous polyposis and familial colon cancer
rb	Retinoblastoma
p53	Li-Fraumeni syndrome, which is characterized by a wide spectrum of tumors, including soft-tissue and bone sarcomas, brain tumors, adrenocortical tumors, and premenopausal breast cancers
Pten	Causes Cowden syndrome, which is characterized by the presence of noncancerous, tumor-like growths called hamartomas in the skin, breast, thyroid, gastrointestinal tract, and central nervous system and results in an increased risk of breast and thyroid carcinomas
BRCA-1	Familial breast cancer
BRCA-2	Familial breast cancer
WT1	Wilms tumor, which is a nephroblastoma
NF1	Neurofibromatosis
MTS1	Hereditary malignant melanoma
MSH2	Nonpolyposis colorectal cancer
MLH1	Nonpolyposis colorectal cancer
XP-A to *XP-G*	UV-sensitive forms of cancer such as basal cell carcinoma
Oncogenes	
RET	Multiple endocrine neoplasia type 2

*Many of the genes described in this table are mutated in more than one type of cancer. The cancers listed are those in which it has been firmly established that a predisposition to develop the disease is commonly due to germ-line mutations in the designated gene.

†*MSH2*, *MLH1*, and *XP-A* to *XP-G* encode proteins that are involved in DNA repair. Some geneticists put these in a separate category from tumor-suppressor genes. However, in this textbook, they are considered tumor-suppressor genes because their loss of function promotes the development of cancer, as shown by the examples in this table.


○, □ Homozygous normal
● Heterozygous for *BRCA-1* mutation, affected
⊙, ⊡ Heterozygous for *BRCA-1* mutation, unaffected
Normal cell
Cancer cell
BRCA-1
BRCA-1
(Inherited mutation)
BRCA-1
(Occurs in a somatic cell)
BRCA-1
(Inherited mutation)
Heterozygote – cell has one functional copy of BRCA-1
Loss of heterozygosity – cell has zero functional copies of BRCA-1

(a) Pedigree for familial breast cancer

(b) Development of cancer at the cellular level

FIGURE 23.19 **Inheritance pattern of familial breast cancer.** **(a)** A family pedigree that involves a loss-of-function mutation in the *BRCA-1* gene. The individuals in this pedigree were tested to determine if they carry a mutation in the *BRCA-1* gene. Affected individuals are shown with closed symbols. The disorder follows a dominant pattern of inheritance. Note: Males can occasionally develop breast cancer though it is much more common in females. **(b)** Breast cancer at the cellular level. Normal cells in an affected individual are heterozygous, whereas cancer cells in the same individual have lost their heterozygosity. Therefore, at the cellular level, cancer is recessive because both alleles must be inactivated for it to occur.

Concept check: Explain why familial breast cancer shows a dominant pattern of inheritance in a pedigree even though it is recessive at the cellular level.

chromosome carrying the normal *BRCA-1* gene could be lost during cell division. Another possible mechanism is mitotic recombination, which is described in Chapter 7. Occasionally, homologous chromosomes can cross over during mitosis. When this occurs in a heterozygote, a daughter cell may receive two copies of the chromosome carrying the normal allele or two copies of the chromosome carrying the mutant allele. The cell carrying two copies of the mutant allele may continue to divide and accumulate additional genetic changes that result in breast cancer.

As noted in Table 23.11, not all hereditary forms of cancer are due to defective tumor-suppressor genes. For example, multiple endocrine neoplasia type 2 is due to the activation of an oncogene.

23.4 REVIEWING THE KEY CONCEPTS

- Cancer is clonal in origin. It is usually a multistep process involving several mutations that begins with benign growth and progresses to growth that is invasive and metastatic (see Figure 23.10).
- Oncogenes often exert their effects via cell division pathways that control the cell cycle (see Figure 23.11).
- Oncogenes are due to gain-of-function mutations in proto-oncogenes. An example is a mutation in the *ras* gene that prevents the Ras protein from hydrolyzing GTP (see Table 23.7, Figure 23.12).
- Different types of genetic changes can convert proto-oncogenes into oncogenes. These include missense mutations, gene amplification, chromosomal translocation, and viral integration (see Table 23.8, Figure 23.13).
- Certain viruses cause cancer by carrying oncogenes into cells (see Table 23.9).
- Knudson's two-hit model proposes that two mutations are needed to cause retinoblastoma (see Figure 23.14).
- The vertebrate *p53* gene senses DNA damage and prevents damaged cells from dividing. It may result in apoptosis (see Figure 23.15).
- Most tumor-suppressor genes encode proteins that are negative regulators of cell division or play a role in genome maintenance (see Table 23.10).
- Checkpoint proteins are a category of proteins that prevent cells that may harbor genetic abnormalities from progressing through the cell cycle (see Figure 23.16).
- Tumor-suppressor genes can be silenced by loss-of-function mutations, chromosome loss, and DNA methylation.
- Most forms of cancer involve multiple genetic changes (see Figure 23.17).
- Alterations in chromosome structure and number are common in cancer cells (see Figure 23.18).
- Molecular profiling is used to classify tumors, which may affect treatment options.
- An inherited predisposition to develop cancer is usually caused by mutations in tumor-suppressor genes (see Table 23.11).

- The predisposition for developing cancer is inherited in a dominant fashion because a heterozygote exhibits the predisposition. At the cellular level, loss of heterozygosity promotes cancer (see Figure 23.19).

23.4 COMPREHENSION QUESTIONS

1. An oncogene is produced from a _________ that has acquired a ____________.
   a. proto-oncogene, loss-of-function mutation
   b. proto-oncogene, gain-of-function mutation
   c. tumor-suppressor gene, loss-of-function mutation
   d. tumor-suppressor gene, gain-of-function mutation
2. Which of the following is a type of genetic change that could produce an oncogene?
   a. Missense mutation
   b. Gene amplification
   c. Chromosomal translocation
   d. All of the above
3. Tumor-suppressor genes promote cancer when
   a. they are overexpressed.
   b. they are expressed in the wrong cell type.
   c. their function is inactivated.
   d. they are expressed at the wrong stage of development.
4. Normal (nonmutant) tumor-suppressor genes often function
   a. as negative regulators of cell division.
   b. in the maintenance of genome integrity.
   c. in the stimulation of cell division.
   d. both a and b.
5. Most forms of cancer involve
   a. the activation of a single oncogene.
   b. the inactivation of a single tumor-suppressor gene.
   c. the activation of multiple oncogenes.
   d. the activation of multiple oncogenes and the inactivation of multiple tumor-suppressor genes.

KEY TERMS

Page 552. monozygotic (MZ) twins, dizygotic (DZ) twins, concordance, age of onset
Page 554. loss-of-function mutations
Page 555. haploinsufficiency, gain-of-function mutations, dominant-negative mutations
Page 556. locus heterogeneity
Page 559. haplotype, haplotype association study, founder
Page 560. linkage disequilibrium
Page 561. genetic testing, genetic screening
Page 562. amniocentesis, chorionic villus sampling (CVS)
Page 563. preimplantation genetic diagnosis (PGD), in vitro fertilization (IVF), prion
Page 565. cancer, clonal, benign, malignant, invasive, metastatic, carcinogen
Page 566. growth factors, oncogene, proto-oncogene
Page 569. acutely transforming viruses (ACTs)
Page 570. tumor-suppressor genes
Page 571. apoptosis, caspases
Page 572. genome maintenance, checkpoint proteins
Page 575. molecular profiling
Page 576. personalized medicine, loss of heterozygosity (LOH)

CHAPTER SUMMARY

- Approximately 12,000 genetic diseases are known to afflict people.

23.1 Inheritance Patterns of Genetic Diseases

- A genetic basis for a human disease may be suggested by a variety of different observations (see Figure 23.1).
- Thousands of human genetic diseases follow a simple Mendelian pattern of inheritance. These include autosomal recessive, autosomal dominant, X-linked recessive, and X-linked dominant (see Figures 23.2–23.4, Tables 23.1–23.4).
- Recessive diseases are usually caused by loss-of-function mutations, whereas dominant diseases may be caused by haploinsufficiency, gain-of-function mutations, or dominant-negative mutations.
- Many genetic diseases exhibit locus heterogeneity, which means that mutations in more than one gene can cause the disorder.

23.2 Detection of Disease-Causing Alleles

- The term *haplotype* refers to the linkage of alleles or molecular markers along a single chromosome (see Figure 23.5).
- Disease-causing mutations may occur in a founder with a specific haplotype (see Figure 23.6).
- Researchers may identify disease-causing alleles by determining their location within a chromosome via haplotype association studies (see Figure 23.7).
- Genetic testing for abnormalities can be performed in a variety of ways (see Table 23.5).
- Genetic screening is population-wide genetic testing.
- Genetic testing can be done prior to birth, which includes amniocentesis, chorionic villus sampling, and preimplantation genetic diagnosis (see Figure 23.8).

23.3 Prions

- Prions are infectious agents composed entirely of protein. They can be acquired by eating infected meat or via a genetic predisposition (see Table 23.6).
- Prions cause disease because the PrP protein in the abnormal (prion) conformation is able to convert normal proteins to the abnormal conformation (see Figure 23.9).

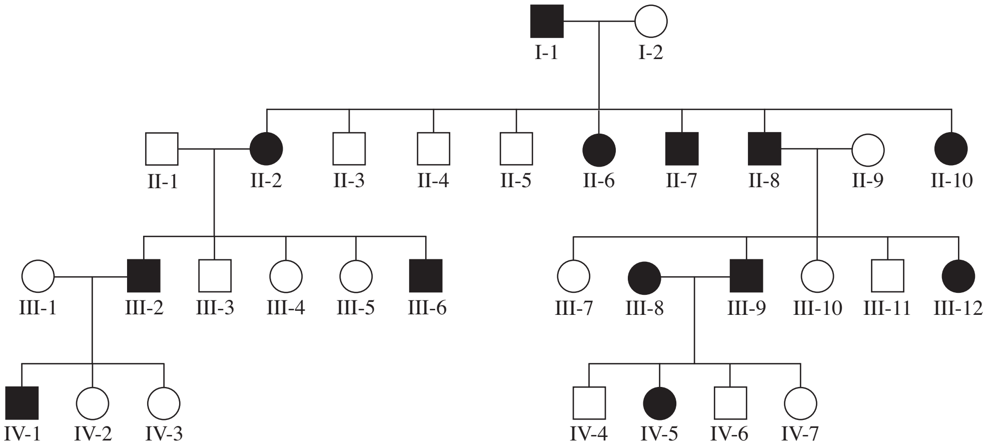
23.4 Genetic Basis of Cancer

- Cancer is clonal in origin. It is usually a multistep process involving several mutations that begins with benign growth and progresses to growth that is invasive and metastatic (see Figure 23.10).
- Oncogenes often exert their effects via cell division pathways that control the cell cycle (see Figure 23.11).
- Oncogenes are due to gain-of-function mutations in proto-oncogenes. An example is a mutation in the *ras* gene that prevents the Ras protein from hydrolyzing GTP (see Table 23.7, Figure 23.12).
- Different types of genetic changes can convert proto-oncogenes into oncogenes. These include missense mutations, gene amplification, chromosomal translocation, and viral integration (see Table 23.8, Figure 23.13).
- Certain viruses cause cancer by carrying oncogenes into cells (see Table 23.9).
- Knudson's two-hit model proposes that two mutations are needed to cause retinoblastoma (see Figure 23.14).
- The vertebrate *p53* gene senses DNA damage and prevents damaged cells from dividing. It may result in apoptosis (see Figure 23.15).
- Most tumor-suppressor genes encode proteins that are negative regulators of cell division or play a role in genome maintenance (see Table 23.10).
- Checkpoint proteins are a category of proteins that prevent cells that may harbor genetic abnormalities from progressing through the cell cycle (see Figure 23.16).
- Tumor-suppressor genes can be silenced by loss-of-function mutations, chromosome loss, and DNA methylation.
- Most forms of cancer involve multiple genetic changes (see Figure 23.17).
- Alterations in chromosome structure and number are common in cancer cells (see Figure 23.18).
- Molecular profiling is used to classify tumors, which may affect treatment options.
- An inherited predisposition to develop cancer is usually caused by mutations in tumor-suppressor genes (see Table 23.11).
- The predisposition for developing cancer is inherited in a dominant fashion because a heterozygote exhibits the predisposition. At the cellular level, loss of heterozygosity promotes cancer (see Figure 23.19).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. The pedigree shown below concerns a human disease known as familial hypercholesterolemia.



This disorder is characterized by an elevation of serum cholesterol in the blood. Though relatively rare, this genetic abnormality can be a contributing factor to heart attacks. At the molecular level, this disease is caused by a mutant gene that encodes a protein called the low-density lipoprotein receptor (LDLR). In the bloodstream, serum cholesterol is bound to a carrier protein known as low-density lipoprotein (LDL). LDL binds to LDLR so that cells can absorb cholesterol. When LDLR is defective, it becomes more difficult for the cells to absorb cholesterol. This explains why the levels of LDL blood cholesterol remain high. Based on the pedigree, what is the most likely pattern of inheritance of this disorder?

Answer: The pedigree is consistent with a dominant pattern of inheritance. An affected individual always has an affected parent. Also, individuals III-8 and III-9, who are both affected, produced unaffected offspring. If this trait was recessive, two affected parents should always produce affected offspring. However, because the trait is dominant, two heterozygous parents can produce homozygous unaffected offspring. The ability of two affected parents to have unaffected offspring is a striking characteristic of dominant inheritance. On average, we would expect that two heterozygous parents should produce 25% unaffected offspring. In the family containing IV-4, IV-5, IV-6, and IV-7, three out of four offspring are unaffected. This higher-than-expected proportion of unaffected offspring is not too surprising because the family is a very small group and may deviate substantially from the expected value due to random sampling error.

S2. We often speak of diseases such as phenylketonuria (PKU) and achondroplasia as having a "genetic basis." Explain whether the following statements are accurate with regard to the genetic basis of any human disease (not just PKU and achondroplasia).

A. An individual must inherit two copies of a mutant allele to have disease symptoms.

B. A genetic predisposition means that an individual has inherited one or more alleles that make it more likely that he or she will develop disease symptoms than other individuals in the population.

C. A genetic predisposition to develop a disease may be passed from parents to offspring.

D. The genetic basis for a disease is always more important than the environment.

Answer:

A. False. Dominant alleles require the inheritance of only one copy, and X-linked recessive alleles require the inheritance of only one copy in males.

B. True.

C. True.

D. False. In many cases, it is difficult to assess the relative contributions of genetics and environment. In some cases, however, the environment seems most important. For example, with PKU, a low-phenylalanine diet prevents harmful symptoms, even if an individual is homozygous for the mutant PKU alleles.

S3. Achondroplasia is a rare form of dwarfism caused by an autosomal dominant mutation that affects the gene that encodes a fibroblast growth factor receptor. Among 1,422,000 live births, the number of babies born with achondroplasia was 31. Among those 31 babies, 18 of them had at least one parent with achondroplasia. The remaining babies had two unaffected parents. How do you explain these 13 babies, assuming that the mutant allele has 100% penetrance? What are the odds that these 13 individuals will pass this mutant gene to their offspring?

Answer: The 13 babies have acquired a new mutation. In other words, during spermatogenesis or oogenesis, or after the egg was fertilized, a new mutation occurred in the fibroblast growth factor gene. These 13 individuals have the same chances of passing the mutant allele to their offspring as the 18 individuals who inherited the mutant allele from a parent. The chance is 50%.

S4. Relatively few inherited forms of cancer involve the inheritance of mutant oncogenes. Instead, most inherited forms of cancer are defects in tumor-suppressor genes. Give two or more reasons why we seldom see inherited forms of cancer involving activated oncogenes.

Answer: If an oncogene was inherited, it may cause uncontrollable cell growth at an early stage of development, thereby causing an embryo to develop improperly. This could lead to an early spontaneous abortion, which would explain why we do not observe individuals with inherited oncogenes. Another possibility is that inherited oncogenes may adversely affect gamete survival, which would make it difficult for them to be passed from parent to offspring. A third possibility would be that oncogenes could affect the fertilized zygote in a way that prevents the successful implantation in the uterus.

S5. Oncogenes sometimes result from genetic rearrangements (e.g., translocations) that produce gene fusions. An example is the Philadelphia chromosome, in which a reciprocal translocation between chromosomes 9 and 22 leads to fusion of the first part of the *bcr* gene with the *abl* gene. Suggest two different reasons why a gene fusion could create an oncogene.

Answer: An oncogene is derived from a genetic change that abnormally activates the expression of a gene that plays a role in cell division. A genetic change that creates a gene fusion can abnormally activate the expression of the gene in two ways.

The first way is at the level of transcription. The promoter and part of the coding sequence of one gene may become fused with the coding sequence of another gene. For example, the promoter and part of the coding sequence of the *bcr* gene may fuse with the coding sequence of the *abl* gene. After this has occurred, the *abl* gene is now under the control of the *bcr* promoter, rather than its own normal promoter. Because the *bcr* promoter is turned on in different cells compared to the *abl* promoter, overexpression of the ABL protein occurs in certain cell types compared with its normal level of expression.

A second way that a gene fusion can cause abnormal activation is at the level of protein structure. A fusion protein has parts of two different polypeptides. The first portion of a fusion protein may affect the structure of the second portion of the polypeptide in such a way that the second portion becomes abnormally active, or vice versa.

Conceptual Questions

C1. For a rare genetic disease that is equally probable in males and females, the concordance for identical twins is 91% and for fraternal twins it is 45%. Do these values indicate a recessive or dominant pattern of inheritance? Suggest two reasons why the concordance value between identical twins is not 100%.

C2. With regard to pedigree analysis, make a list of the patterns that distinguish recessive, dominant, and X-linked genetic diseases from each other.

C3. Explain, at the molecular level, why human genetic diseases often follow a simple Mendelian pattern of inheritance, whereas most normal traits, such as the shape of your nose or the size of your head, are governed by multiple gene interactions.

C4. Many genetic disorders exhibit locus heterogeneity. Define and give two examples of locus heterogeneity. How does locus heterogeneity confound a pedigree analysis?

C5. In general, why do changes in chromosome structure or number tend to affect an individual's phenotype? Explain why some changes in chromosome structure, such as reciprocal translocations, do not.

C6. Figure 23.1 illustrates albinism in different species. Describe two other genetic disorders found in both humans and animals.

C7. Discuss why a genetic disease might have a particular age of onset. Would an infectious disease have an age of onset? Explain why or why not.

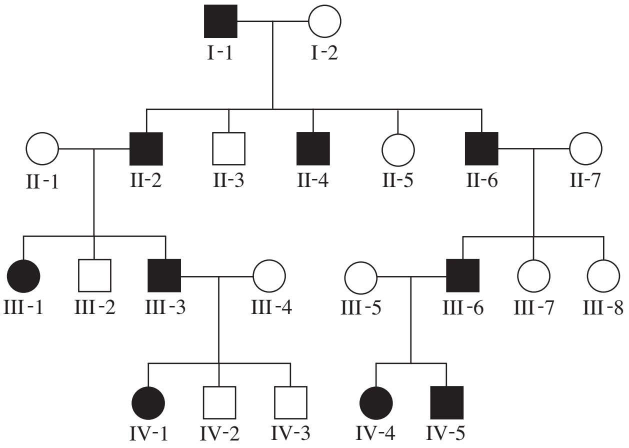
C8. Gaucher disease (type I) is due to a defect in a gene that encodes a protein called acid β-glucosidase. This enzyme plays a role in carbohydrate metabolism within the lysosome. The gene is located on the long arm of chromosome 1. People who inherit two defective copies of this gene exhibit Gaucher disease, the major symptoms of which include an enlarged spleen, bone lesions, and changes in skin pigmentation. Let's suppose a phenotypically unaffected woman, whose father had Gaucher disease, has a child with a phenotypically unaffected man, whose mother had Gaucher disease.

A. What is the probability that this child will have the disease?

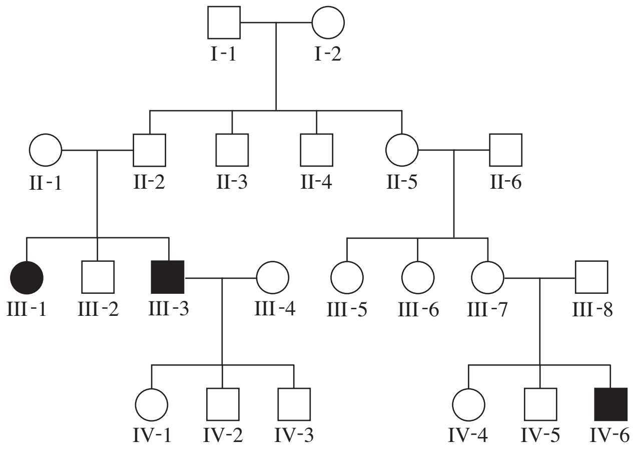
B. What is the probability that this child will have two normal copies of this gene?

C. If this couple has five children, what is the probability that one of them will have Gaucher disease and four will be phenotypically unaffected?

C9. Ehler-Danlos syndrome is a relatively rare disorder caused by a mutation in a gene that encodes a protein called collagen (type 3 A1). Collagen is a protein found in the extracellular matrix that plays an important role in the formation of skin, joints, and other connective tissues. People with this syndrome have extraordinarily flexible skin and very loose joints. The pedigree shown below contains several members affected with Ehler-Danlos syndrome, shown with black symbols. Based on this pedigree, does this syndrome appear to be an autosomal recessive, autosomal dominant, X-linked recessive, or X-linked dominant trait? Explain your reasoning.

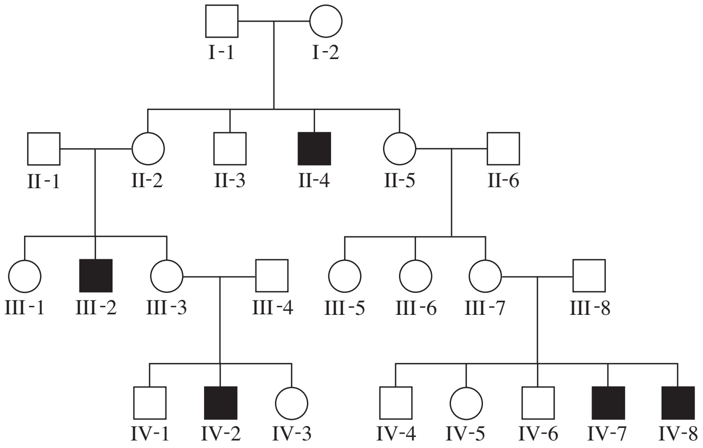


C10. Hurler syndrome is due to a mutation in a gene that encodes a protein called α-L-iduronidase. This protein functions within the lysosome as an enzyme that breaks down mucopolysaccharides (a type of polysaccharide that has many acidic groups attached). When this enzyme is defective, excessive amounts of the mucopolysaccharides dermatan sulfate and heparin sulfate accumulate within the lysosomes, especially in liver cells and connective tissue cells. This leads to symptoms such as an enlarged liver and spleen, bone abnormalities, corneal clouding, heart problems, and severe neurological problems. The pedigree shown here contains three members affected with Hurler syndrome, indicated with black symbols. Based on this pedigree, does this syndrome appear to be an autosomal recessive, autosomal dominant, X-linked recessive, or X-linked dominant trait? Explain your reasoning.



C11. Like Hurler syndrome, Fabry disease involves an abnormal accumulation of substances within lysosomes. However, the lysosomes of individuals with Fabry disease show an abnormal accumulation of lipids. The defective enzyme is α-galactosidase A, which is a lysosomal enzyme that functions in lipid metabolism. This defect causes cell damage, especially to the kidneys, heart, and eyes. The gene that encodes α-galactosidase A is found on the X chromosome. Let's suppose a phenotypically unaffected couple produces two sons with Fabry disease and one phenotypically unaffected daughter. What is the probability that the daughter will have an affected son?

C12. Lesch-Nyhan syndrome is due to a mutation in a gene that encodes a protein called hypoxanthine-guanine phosphoribosyltransferase (HPRT). HPRT is an enzyme that functions in purine metabolism. People afflicted with this syndrome have severe neurodegeneration and loss of motor control. The pedigree shown below contains several members with Lesch-Nyhan syndrome. Affected members are shown with black symbols. Based on this pedigree, does this syndrome appear to be an autosomal recessive, autosomal dominant, X-linked recessive, or X-linked dominant trait? Explain your reasoning.

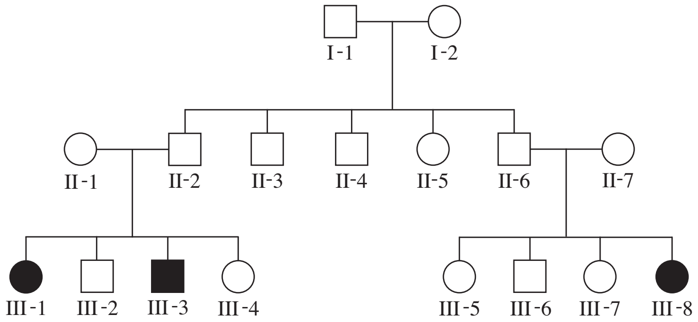


C13. Marfan syndrome is due to a mutation in a gene that encodes a protein called fibrillin-1. It is inherited as a dominant trait. The fibrillin-1 protein is the main constituent of extracellular microfibrils. These microfibrils can exist as individual fibers or associate with a protein called elastin to form elastic fibers. The gene that encodes fibrillin-1 is located on the long arm of chromosome 15. Let's suppose a phenotypically unaffected woman has a child with a man who has Marfan syndrome.

A. What is the probability this child will have the disease?

B. If this couple has three children, what is the probability none of them will have Marfan syndrome?

C14. Sandhoff disease is due to a mutation in a gene that encodes a protein called hexosaminidase B. This disease has symptoms that are similar to Tay-Sachs disease. Weakness begins in the first 6 months of life. Individuals exhibit early blindness and progressive mental and motor deterioration. The pedigree shown here contains three members with Sandhoff disease. Affected members are shown with black symbols.



A. Based on this pedigree, does this syndrome appear to be an autosomal recessive, autosomal dominant, X-linked recessive, or X-linked dominant trait? Explain your reasoning.

B. What is the likelihood that II-1, II-2, II-3, II-4, II-5, II-6, and II-7 carry a mutant allele for the hexosaminidase B gene?

C15. What is a haplotype? Describe the rationale of a haplotype association study to locate a disease-causing allele.

C16. What is a prion? Explain how a prion relies on normal cellular proteins to cause a disease such as mad cow disease.

C17. Some people have a genetic predisposition for developing prion diseases. Examples are described in Table 23.6. In the case of Gerstmann-Straüssler-Scheinker disease, the age of onset is typically at 30 to 50 years, and the duration of the disease (which leads to death) is about 5 years. Suggest a possible explanation why someone can live for a relatively long time without symptoms and then succumb to the disease in a relatively short time.

C18. Familial fatal insomnia (described in Table 23.6) is a prion disease inherited as an autosomal dominant trait. Researchers have identified the *PrP* gene, located on the short arm of chromosome 20, and tried to understand the relationship between the *PrP* gene sequence and the molecular basis of familial fatal insomnia.

A. It has been found that a rare mutation at codon 178, changing an aspartic acid to an asparagine, can cause this disease. In addition, codon 129 seems to play a role. The human population is polymorphic at codon 129, which may encode valine or methionine. If codon 178 is an aspartic acid (Asp) codon, it does not seem to matter if valine or methionine is found at position 129. In other words, if codon 178 is aspartic acid, Met-129 and Val-129 codons are not associated with disease symptoms. However, if codon 178 specifies asparagine (Asn), then it does matter. Familial fatal insomnia seems to require an asparagine at codon 178 and a methionine at codon 129 in the *PrP* gene sequence. Although the actual biochemical mechanism is not known, suggest a possible reason why this is the case.

B. Also, researchers have compared the sequences of the *PrP* gene in many people with familial fatal insomnia. Keep in mind that this is a dominant autosomal trait, so people with this disorder have one mutant copy of the *PrP* gene and one normal copy. People with familial fatal insomnia must have one abnormal copy of the gene that contains Asn-178 and Met-129. The second copy of the gene has Asp-178, and it may have Met-129 or Val-129. Some results suggest that people having a Met-129 codon in this second copy of the *PrP* gene causes the disease to develop more rapidly than in people who have Val-129 in the second copy. Propose an explanation why disease symptoms may occur more rapidly when Met-129 is found in the second copy of the *PrP* gene.

C19. What is the difference between an oncogene and a tumor-suppressor gene? Give two examples.

C20. What is a proto-oncogene? What are the typical functions of proteins encoded by proto-oncogenes? At the level of protein function, what are the general ways that proto-oncogenes can be converted into oncogenes?

C21. What is a retroviral oncogene? Is it necessary for viral infection and proliferation? How have retroviruses acquired oncogenes?

C22. A genetic predisposition to developing cancer is usually inherited as a dominant trait. At the level of cellular function, are the alleles involved actually dominant? Explain why some individuals who have inherited these dominant alleles do not develop cancer during their lifetimes.

C23. Describe the types of genetic changes that commonly convert a proto-oncogene into an oncogene. Give three examples. Explain how the genetic changes are expected to alter the activity of the gene product.

C24. The *rb* gene encodes a protein that inhibits E2F, a transcription factor that activates several genes involved in cell division. Mutations in *rb* are associated with certain forms of cancer, such as retinoblastoma. Under each of the following conditions, would you expect cancer to occur?

A. One copy of *rb* is defective; both copies of *E2F* are functional.

B. Both copies of *rb* are defective; both copies of *E2F* are functional.

C. Both copies of *rb* are defective; one copy of *E2F* is defective.

D. Both copies of *rb* and *E2F* are defective.

C25. A *p53* knockout mouse in which both copies of *p53* are defective has been produced by researchers. This type of mouse appears normal at birth. However, it is highly sensitive to UV light. Based on your knowledge of *p53*, explain the normal appearance at birth and the high sensitivity to UV light.

C26. With regard to cancer cells, which of the following statements are true?

A. Cancer cells are clonal, which means they are derived from a single mutant cell.

B. To become cancerous, cells usually accumulate multiple genetic changes that eventually result in uncontrolled growth.

C. Most cancers are caused by oncogenic viruses.

D. Cancer cells have lost the ability to properly regulate cell division.

C27. When the DNA of a human cell becomes damaged, the *p53* gene is activated. What is the general function of the p53 protein? Is it an enzyme, transcription factor, cell cycle protein, or something else? Describe three ways that the synthesis of the p53 protein affects cellular function. Why is it beneficial for these three things to happen when a cell's DNA has been damaged?

Application and Experimental Questions

E1. Which of the following experimental observations would suggest that a disease has a genetic basis?

A. The frequency of the disease is less likely in relatives that live apart compared to relatives that live together.

B. The frequency of the disease is unusually high in a small group of genetically related individuals who live in southern Spain.

C. The disease symptoms usually begin around the age of 40.

D. It is more likely that both monozygotic twins will be affected by the disease compared to dizygotic twins.

E2. At the beginning this chapter, we discussed the types of experimental observations that suggest a disease is inherited. Which of these observations do you find the least convincing? Which do you find the most convincing? Explain your answer.

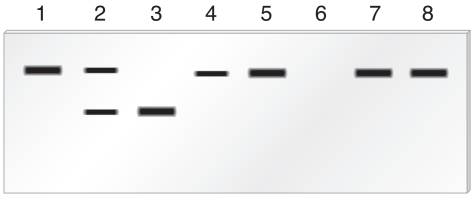
E3. What does the term *genetic testing* mean? What is the difference between testing at the protein level and testing at the DNA level? Describe five different techniques used in genetic testing.

E4. A particular disease is found in a group of South American Indians. During the 1920s, many of these people migrated to Central America. In the Central American group, the disease is never found. Discuss whether or not you think the disease has a genetic component. What types of further observations would you make?

E5. Chapter 19 describes a blotting method known as Western blotting that can be used to detect a polypeptide that is translated from a particular mRNA. In this method, a particular polypeptide or protein is detected by an antibody that specifically recognizes a segment of its amino acid sequence. After the antibody binds to the polypeptide within a gel, a secondary antibody (which is labeled) is used to visualize the polypeptide as a dark band. For example, an antibody that recognizes α-galactosidase A could be used to specifically detect the amount of α-galactosidase A protein on a gel. The enzyme α-galactosidase A is defective in individuals with Fabry disease, which shows an X-linked recessive pattern of inheritance. Amy, Nan, and Pete are siblings, and Pete has Fabry disease. Aileen, Jason, and Jerry are brothers and sister, and Jerry has Fabry disease. Amy, Nan, and Pete are not related to Aileen, Jason, and Jerry. Amy, Nan, and Aileen are concerned they could be carriers of a defective α-galactosidase A gene. A sample of cells was obtained from each of these six individuals and subjected to Western blotting, using an antibody against α-galactosidase A. Samples were also obtained from two unrelated normal individuals (lanes 7 and 8). The results are shown here.

Samples from:

Lane 1. Amy
Lane 2. Nan
Lane 3. Pete
Lane 4. Aileen
Lane 5. Jason
Lane 6. Jerry
Lane 7. Normal male
Lane 8. Normal female

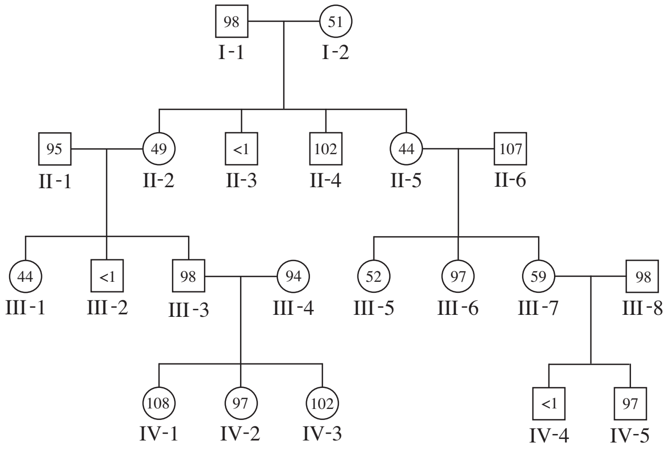


Note: Due to X inactivation in females, the amount of expression of genes on the single X chromosome in males is equal to the amount of expression from genes on both X chromosomes in females.

A. Explain the type of mutation (i.e., missense, nonsense, promoter, etc.) that causes Fabry disease in Pete and Jerry.

B. What would you tell Amy, Nan, and Aileen regarding the likelihood they are carriers of the mutant allele and the probability of having affected offspring?

E6. An experimental assay for the blood clotting protein called factor IX is available. A blood sample was obtained from each member of the pedigree shown here. The amount of factor IX protein is shown within the symbol of each member and is expressed as a percent of the amount observed in normal individuals who do not carry a mutant copy of the gene.



What are the likely genotypes of each member of this pedigree?

E7. Explain how the experimental study of cancer-causing viruses has increased our understanding of cancer.

E8. Discuss ways to distinguish whether a particular form of cancer involves an inherited predisposition or is due strictly to (postzygotic) somatic mutations. In your answer, consider that only one mutation may be inherited, but the cancer might develop only after several somatic mutations.

E9. The codon change (Gly-12 to Val-12) in human *ras*H that converts it to oncogenic *ras*H has been associated with many types of cancers. For this reason, researchers would like to develop drugs to inhibit oncogenic *ras*H. Based on your understanding of the Ras protein, what types of drugs might you develop? In other words, what would be the structure of the drugs, and how would they inhibit the Ras protein? How would you test the efficacy of the drugs? What might be some side effects?

Questions for Student Discussion/Collaboration

1. Make a list of the benefits that may arise from genetic testing as well as possible negative consequences. Discuss the items on your list.
2. Our government has finite funds to devote to cancer research. Discuss which aspects of cancer biology you would spend the most money pursuing.
   A. Identifying and characterizing oncogenes and tumor-suppressor genes
   B. Identifying agents in our environment that cause cancer
   C. Identifying viruses that cause cancer
   D. Devising methods aimed at killing cancer cells in the body
   E. Informing the public of the risks involved in exposure to carcinogens

   In the long run, which of these areas would you expect to be the most effective in decreasing mortality due to human cancer?

Answers to Comprehension Questions

23.1: b, a, b, b

23.2: c, c, a, c

23.3: c

23.4: b, d, c, d, d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

24

The fruit fly (Drosophila melanogaster), a model organism in the study of developmental genetics.

CHAPTER OUTLINE

24.1 Overview of Animal Development

24.2 Invertebrate Development

24.3 Vertebrate Development

24.4 Plant Development

24.5 Immunogenetics

DEVELOPMENTAL GENETICS AND IMMUNOGENETICS

Multicellular organisms, such as animals and plants, begin their lives with a fairly simple organization, namely, a fertilized egg, and then proceed step by step to a much more complex arrangement. As this occurs, cells divide and change their characteristics as they become highly specialized units within a multicellular individual. Each cell in an organism has its own particular role. In animals, for example, muscle cells allow an organism to move, and intestinal cells facilitate the absorption of nutrients. This division of labor among the various cells and organs of the body promotes the survival of the individual.

Developmental genetics is concerned with the roles that genes play in orchestrating the changes that occur during development. In this chapter, we will examine how the sequential actions of genes provide a program for the development of an organism from a fertilized egg to an adult. The last couple of decades have seen staggering advances in our understanding of developmental genetics at the molecular level. Scientists have chosen a few model organisms, such as the fruit fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), mouse (*Mus musculus*), zebrafish (*Danio rerio*) and a flowering plant (*Arabidopsis thaliana*), and worked toward the identification and characterization of the genes required for running their developmental programs. The study of these model organisms has identified general themes in development that apply more broadly to all species.

In certain organisms, notably *Drosophila*, many of the genes that play a critical role in the early stages of development have been identified. Researchers are now studying how the proteins encoded by these genes control the course of development. This research also helps us appreciate the development of humans. In this chapter, we will explore the body plans of invertebrates, vertebrates, and plants, and consider several examples in which geneticists understand how the actions of genes govern the developmental process. We will also explore how genes that function within the immune systems of vertebrates undergo genetic changes that enable them to produce a great diversity of antibodies to fight foreign invaders.

24.1 OVERVIEW OF ANIMAL DEVELOPMENT

Although the details differ widely among various species, the general features and steps in animal development are largely conserved within this kingdom. In this section, we will examine the types of molecules that promote developmental changes and consider how genes play an underlying role in orchestrating a plan of development from a fertilized egg to the adult.

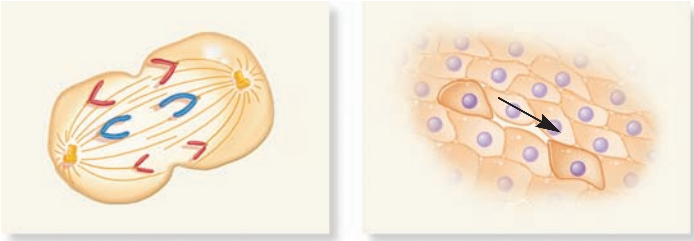
The Generation of a Body Pattern Depends on the Positional Information That Each Cell Receives During Development

Multicellular development in animals (and plants) follows a body plan, or pattern. The term **pattern** refers to the spatial arrangement of different regions of the body. At the cellular level, the body pattern is due to the arrangement of cells and their specialization. For an organism to develop a body pattern with unique morphological and cellular features, each cell of the body must receive **positional information**—signals that cause a cell to follow a particular developmental pathway based on its position relative to the other cells.

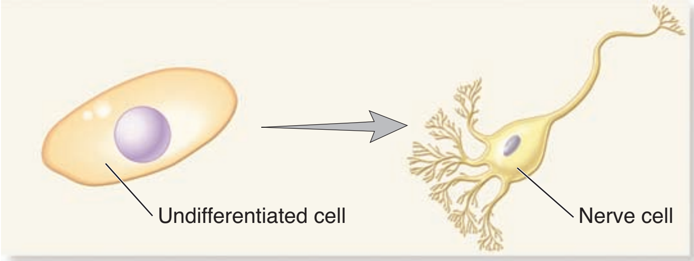
How may a cell respond to positional information? Four basic responses are possible: cell division, cell migration, cell differentiation, and cell death (**Figure 24.1**). For example, positional information may cause a cell or group of cells to divide into daughter cells or migrate in a particular direction from one region of the embryo to another. In addition, positional information may stimulate a cell to differentiate into a particular cell type. Finally, positional information or other stimuli may cause a cell to undergo programmed cell death. This process, known as **apoptosis,** is a necessary event during normal development for the sculpting of tissues and organs. For example, in the later stages of development of the retina in *Drosophila*, excess cells are eliminated by apoptosis. During mammalian development, individual fingers and toes are created when embryonic cells that lie between each future finger and toe undergo this type of cell death. The coordination of these four events leads to the formation of a body with a particular pattern. As we will see, the temporal expression of genes and the localization of gene products at precise regions in the fertilized egg and early embryo are the critical phenomena that underlie this coordination.

Morphogens and Cell Adhesion Molecules Convey Positional Information

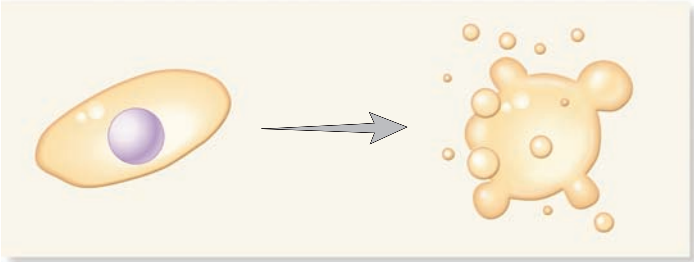
A **morphogen** is a molecule that conveys positional information and promotes developmental changes over time. It is generally a diffusible molecule that acts in a concentration-dependent manner to influence the developmental fate of a cell. Morphogens provide the positional information that stimulates a cell to divide, migrate, differentiate, or die. For example, when a region of a *Drosophila* embryo is exposed to a high concentration of the morphogen known as Bicoid, it develops structures that are characteristic of the anterior region of the body. Many such morphogens function as transcription factors that regulate the expression of many genes. This topic is described in greater detail later. How do morphogens control pattern development? Within an oocyte and during embryonic development, morphogens typically are distributed along a concentration gradient. In other words, the concentration of a morphogen varies from low to high in different regions of the developing organism. A key feature of morphogens is they act in a concentration-dependent manner. In a particular concentration range, a morphogen or a combination of two or more different morphogens restricts a cell into a specific developmental pathway.



(a) Cell division **(b) Cell migration**



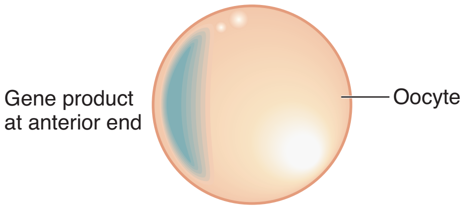
(c) Cell differentiation



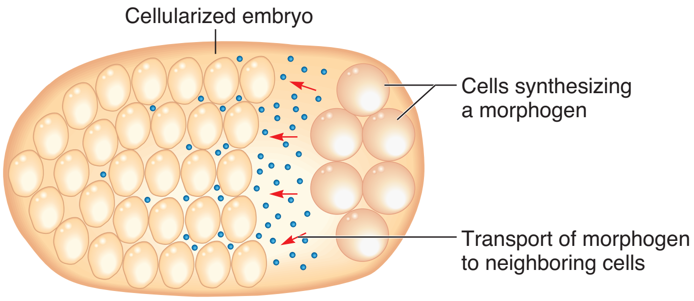
(d) Cell death (apoptosis)

FIGURE 24.1 **Four events that may happen to cells during the course of animal development.**

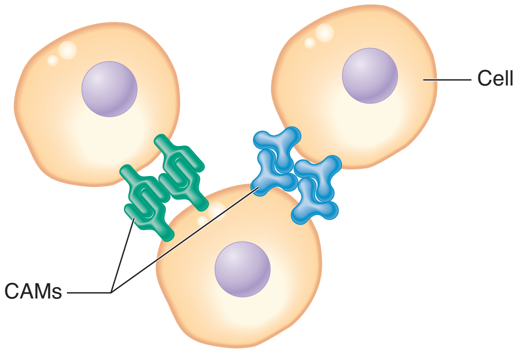
During the earliest stages of development of certain species, morphogenic gradients are preestablished within the oocyte (**Figure 24.2a**). Following fertilization, the zygote subdivides into many smaller cells. Due to the preestablished gradient of morphogens within the oocyte, these smaller cells have higher or lower concentrations of morphogens, depending on their location in the embryo. In this way, the morphogen gradients in the oocyte provide positional information that is important in establishing the general polarity of an embryo along two main axes:



(a) Asymmetrical distribution of morphogens in an oocyte


Cellularized embryo
Cells synthesizing a morphogen
Transport of morphogen to neighboring cells

(b) Asymmetrical synthesis and extracellular distribution of a morphogen in an embryo



(c) Cell-to-cell contact conveys positional information

FIGURE 24.2 **Three molecular mechanisms of positional information.**

Concept check: Which of these mechanisms involves diffusible morphogens?

the anteroposterior axis and the dorsoventral axis, which are described later.

A morphogen gradient can also be established in the embryo by secretion and transport. A certain cell or group of cells may synthesize and secrete a morphogen at a specific stage of development. After secretion, the morphogen is transported to neighboring cells. The concentration of the morphogen is usually highest near the cells that secrete it (Figure 24.2b). The morphogen may then influence the developmental fate of the cells exposed to it. The process by which a cell or group of cells governs the developmental fate of neighboring cells is known as **induction.** An example of a secreted morphogen in *Drosophila* is a protein known as Hedgehog. This protein is secreted from cells on the posterior side of body segments and forms a gradient from the posterior to anterior region of each segment. Hedgehog plays a role in creating the morphological differences between the anterior and posterior parts of individual body segments and also influences the formation of certain adult structures. Hedgehog mutants result in larvae that are short and stubby and have a higher density of denticles, which are cuticle spikes projecting from the body surface. The appearance of the stubby and spikey larvae inspired the name hedgehog.

In addition to morphogens, positional information is conveyed by **cell adhesion** (Figure 24.2c). Each cell makes its own collection of surface receptors that enable it to adhere to other cells and/or to the extracellular matrix (ECM), which consists primarily of carbohydrates and fibrous proteins. Such receptors are known as **cell adhesion molecules (CAMs).** A cell may gain positional information via the combination of contacts it makes with other cells or with the ECM. The contacts, in turn, may activate intracellular signal transduction pathways that lead to developmental changes.

The phenomenon of cell adhesion, and its role in multicellular development, was first recognized by Henry Wilson in 1907. He took multicellular sponges and passed them through a sieve, which disaggregated them into individual cells. Remarkably, the cells actively migrated until they adhered to one another to form a new sponge, complete with the chambers and canals that characterize a sponge's internal structure! When sponge cells from different species were mixed, they sorted themselves properly, adhering only to cells of the same species. Overall, these results indicate that cell adhesion plays an important role in governing the position that a cell adopts during development.

The Study of Mutants with Disrupted Developmental Patterns Has Identified Genes That Control Development

Mutations that alter the course of development in experimental organisms, such as *Drosophila*, have greatly contributed to our understanding of the normal process of development. For example, Figure 24.3 shows a photograph of a fly that carries multiple mutations in a cluster of genes called *bithorax*. This mutant fly has four wings instead of two; the halteres (balancing organs that resemble miniature wings), which are found on the third thoracic segment, are changed into wings, normally found on the second thoracic segment. (The term *bithorax* refers to the observation that the characteristics of the second thoracic segment are duplicated.)

Edward Lewis, a pioneer in the genetic study of development, became interested in the bithorax phenotype and began investigating it in 1946. Researchers later discovered that the mutant chromosomal region actually contained a complex of three genes involved in specifying developmental pathways in the fly. A gene that plays a central role in specifying the final identity of a body region is called a **homeotic gene.** We will discuss particular examples of homeotic genes later in this chapter.

During the 1960s and 1970s, interest in the relationship between genetics and embryology blossomed as biologists began to appreciate the role of genetics at the molecular and cellular

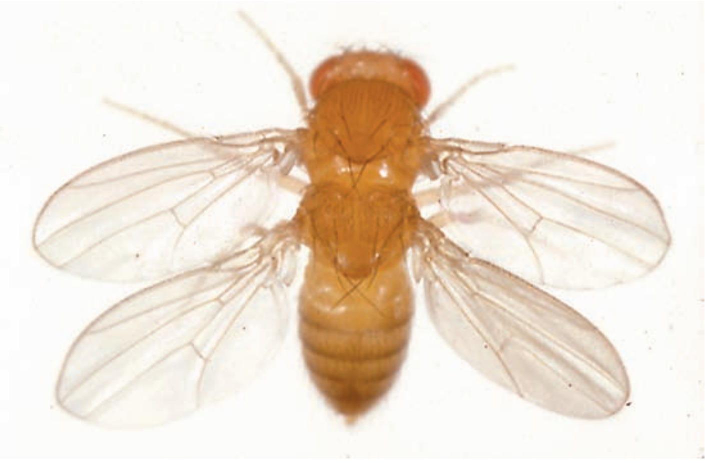


FIGURE 24.3 **The bithorax mutation in *Drosophila*.**

Genes → Traits A normal fly has two wings on the second thoracic segment and two halteres on the third thoracic segment. However, this mutant fly carries multiple mutations in a cluster of genes called the *bithorax* complex. In the mutant fly, the third thoracic segment has the same characteristics as the second thoracic segment, thereby producing a fly with four wings instead of the normal number of two.

levels. It soon became clear that the genomes of multicellular organisms contain groups of genes that initiate a program of development involving networks of gene regulation. By identifying mutant alleles that disrupt development, geneticists have begun to unravel the pattern of gene expression that underlies the normal pattern of multicellular development.

Animal Development Occurs in Four Overlapping Phases

With the exception of sponges, all animal bodies are organized along axes. Some simple animals, such as jellyfish, exhibit radial symmetry in which the body is circularly symmetrical around an axis. However, most animals, including humans, are **bilaterians,** meaning they have left-right symmetry. Our discussion of animal development in this chapter will focus on the bilaterians. For these species, the body is typically organized along two major axes: an anteroposterior axis and a dorsoventral axis. Left-right symmetry occurs relative to the anteroposterior axis.

By comparing a variety of bilaterians, researchers have discovered that development generally proceeds in four overlapping stages (Figure 24.4):

1. ***Formation of body axes:*** The first phase of pattern development is the establishment of the body axes.
2. ***Segmentation of the body:*** In the second phase, the body is subdivided into segments. In invertebrates, the segments may remain morphologically distinct, whereas in vertebrates, distinct segments are obvious only in the early stages of development.
3. ***Determination of structures within the segments:*** As the segments form, groups of cells become destined to develop into particular structures and cell types. This process, which is called **determination,** occurs before the structures and cell types have changed their morphologies.

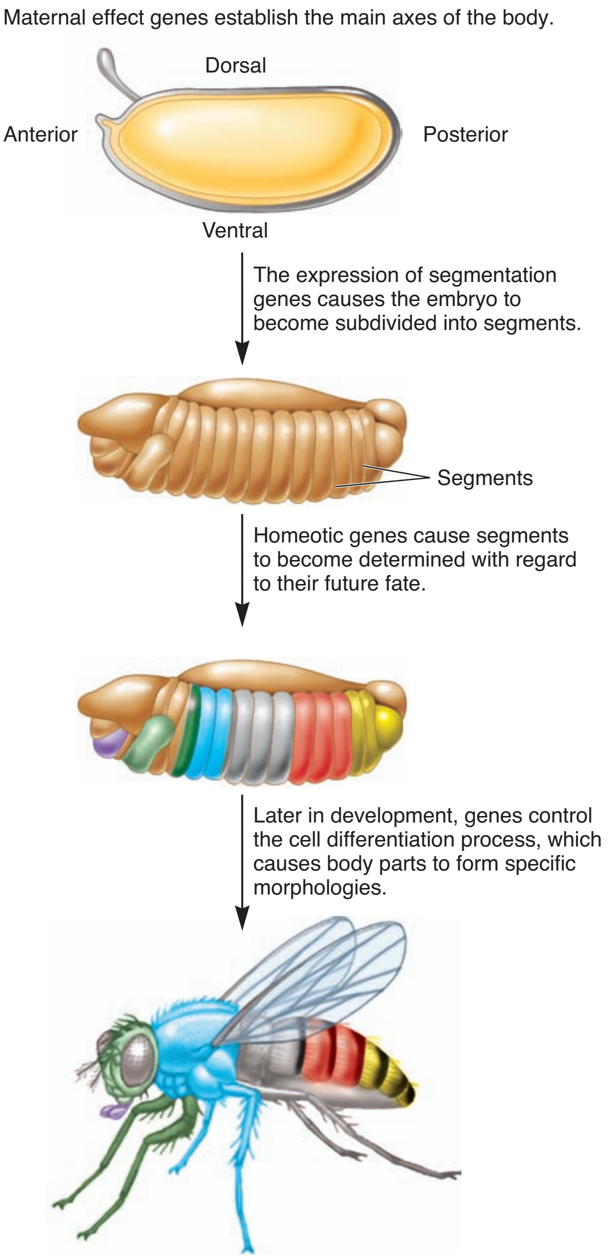

Maternal effect genes establish the main axes of the body.
Dorsal
Anterior
Posterior
Ventral
The expression of segmentation genes causes the embryo to become subdivided into segments.
Segments
Homeotic genes cause segments to become determined with regard to their future fate.
Later in development, genes control the cell differentiation process, which causes body parts to form specific morphologies.

FIGURE 24.4 **The four overlapping phases of animal development.**

Concept check: *How does genetics play a role in development?*

4. ***Cell differentiation:*** Toward the end of development, cells differentiate into particular cell types. This results in tissues and organs with specific morphologies.

The formation of a bilaterian body with a specific pattern is the result of these four overlapping phases. The phases themselves are controlled by sets of genes. One set is expressed very early and controls the formation of body axes. As you will learn later, the genes in this category often exhibit a maternal effect pattern of inheritance. A second set is expressed next that results in segmentation. In *Drosophila*, these are called segmentation genes. A third set, known as homeotic genes, controls the fate of groups of cells within segments. Finally, a fourth set of genes controls the cellular differentiation process. In the following two sections, we will explore the phases of animal development in greater detail and discuss examples of genes that play a key role in these four phases.

24.1 REVIEWING THE KEY CONCEPTS

- Positional information in animals may cause a cell to divide, migrate, differentiate, or undergo apoptosis (see Figure 24.1).
- A morphogen is a molecule that conveys positional information. Three molecular mechanisms that convey positional information include asymmetrical distribution of maternally deposited morphogens in an oocyte, asymmetrical synthesis and extracellular distribution of a morphogen in a developing embryo, and cell-to-cell contact (see Figure 24.2).
- The study of mutants that disrupt development has identified genes that control development (see Figure 24.3).
- Development in animals involves four overlapping phases: the formation of axes, segmentation, determination, and cell differentiation (see Figure 24.4).

24.1 COMPREHENSION QUESTIONS

1. Positional information may provide a cue for a cell to
   a. divide.
   b. migrate.
   c. differentiate.
   d. undergo apoptosis.
   e. do all of the above.
2. Molecules that convey positional information include
   a. morphogens.
   b. cell adhesion molecules.
   c. ATP.
   d. both a and b.
3. Which of the following is the correct order for developmental phases in animals?
   A. Segmentation of the body
   B. Determination
   C. Cell differentiation
   D. Formation of body axes

   a. A, B, C, D
   b. A, D, C, B
   c. D, A, B, C
   d. D, A, C, B

24.2 INVERTEBRATE DEVELOPMENT

We will focus our discussion of invertebrate development by considering two model organisms, the fruit fly (*Drosophila melanogaster*) and a nematode (*Caenorhabditis elegans*), that have been pivotal in our understanding of developmental genetics. As we first saw in Chapter 4, *Drosophila* has been a favorite subject of geneticists since 1910, when Thomas Hunt Morgan isolated his first white-eyed mutant. The fruit fly has been used to discover many of the fundamental principles of genetics, including the chromosome theory of inheritance, the random mutation theory, and linkage mapping, to mention only a few.

In developmental genetics, *Drosophila* is also useful for a variety of reasons. The techniques for generating and analyzing mutants in this organism are highly advanced, and researchers have identified many mutant strains with altered development. Also, *Drosophila* has a long history of genetic and developmental studies, which has been more recently enhanced by the available genome sequence.

By comparison, *C. elegans* is used by developmental geneticists because of its simplicity. The adult organism is a small transparent worm composed of only about 1000 somatic cells. Starting with the fertilized egg, the pattern of cell division and the fate of each cell are completely known.

In this section, we begin by examining the general features of *Drosophila* development. We then focus our attention on embryonic development (embryogenesis) because it is during this stage that the overall body plan is determined. We will see how the expression of particular genes and the localization of gene products within the embryo influence the developmental process. We then briefly consider the pattern of cellular development in *C. elegans*.

The Early Stages of Embryonic Development Determine the Pattern of Structures in the Adult Organism

Figure 24.5 illustrates the general sequence of events in *Drosophila* development. The oocyte is the cell most critical in determining the outcome of development seen in the adult organism. It is an elongated cell with preestablished axes (Figure 24.5a). After fertilization takes place, the zygote goes through a series of nuclear divisions that are not accompanied by cytoplasmic division. Initially, the resulting nuclei are scattered throughout the yolk, but by about 1.5 hours many of them migrate to the periphery of the cytoplasm. This is the syncytial blastoderm stage (Figure 24.5b).

After the nuclei have lined up along the cell membrane, individual cells are formed as portions of the cell membrane surround each nucleus, creating a cellular blastoderm (Figure 24.5c). This structure is composed of a sheet of cells on the outside with yolk in the center. At the posterior end are a group of cells called the pole cells, the primordial germ cells that eventually give rise to gametes in the adult organism. After blastoderm formation is complete, dramatic changes occur during gastrulation (Figure 24.5d). This stage produces three cell layers known as the ectoderm, mesoderm, and endoderm.

As this process occurs, the embryo begins to subdivide into detectable units. Boundaries are formed that divide the embryo into morphologically discrete **segments.** Figure 24.5e shows the

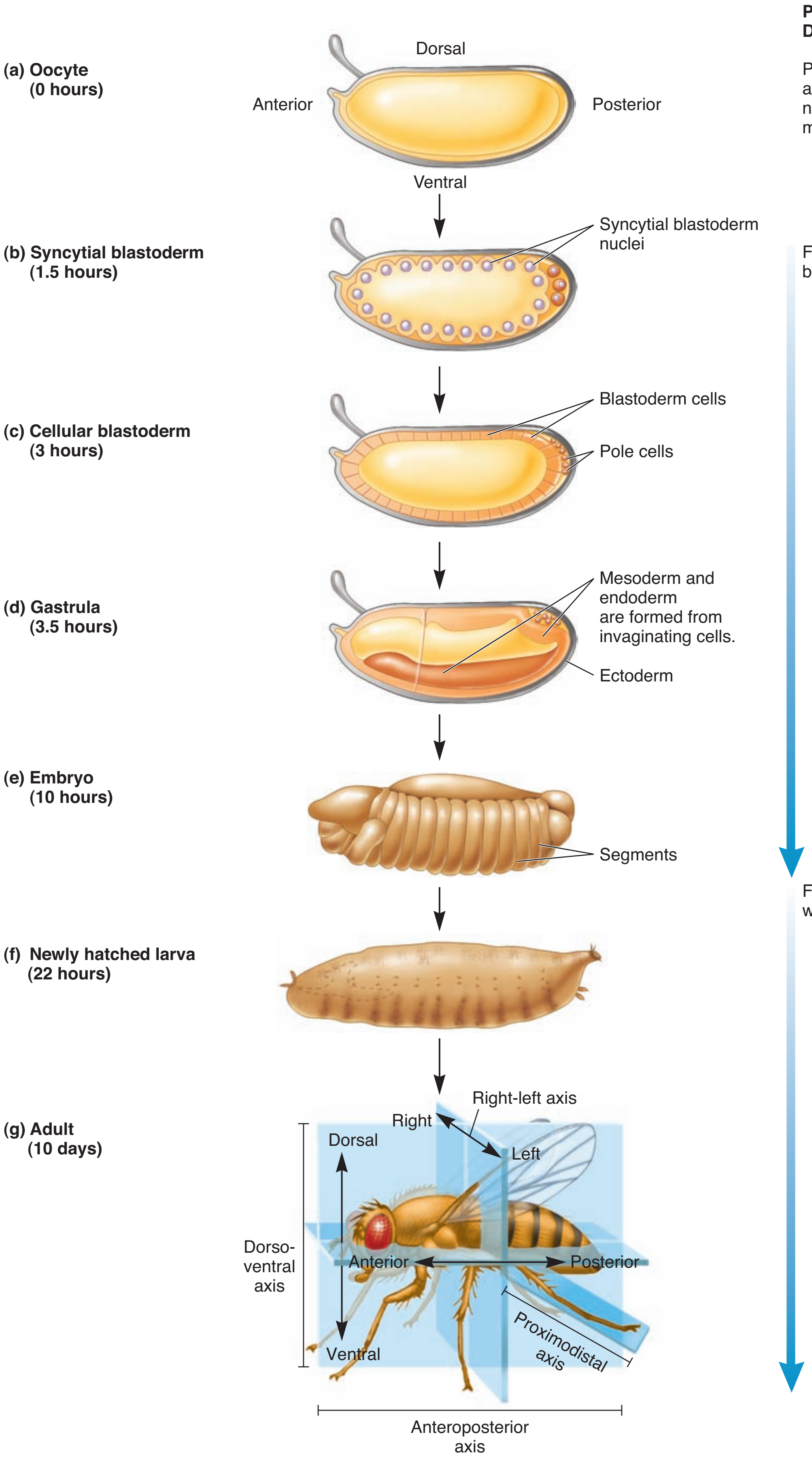

Progression of Developmental Events
(a) Oocyte (0 hours)
Dorsal
Anterior
Posterior
Ventral
Preestablishment of axes in oocyte, by nurse cells in the mother's ovary
(b) Syncytial blastoderm (1.5 hours)
Syncytial blastoderm nuclei
Formation of body segments
(c) Cellular blastoderm (3 hours)
Blastoderm cells
Pole cells
(d) Gastrula (3.5 hours)
Mesoderm and endoderm are formed from invaginating cells.
Ectoderm
(e) Embryo (10 hours)
Segments
Formation of structures within each segment
(f) Newly hatched larva (22 hours)
(g) Adult (10 days)
Right-left axis
Right
Left
Dorsal
Dorso-ventral axis
Anterior
Posterior
Ventral
Proximodistal axis
Anteroposterior axis

FIGURE 24.5 **Developmental stages of the fruit fly *Drosophila*.** **(a)** Oocyte. **(b)** Syncytial blastoderm. **(c)** Cellular blastoderm. **(d)** Gastrula. **(e)** Embryo at 10 hours. **(f)** Newly hatched larva. **(g)** Adult. Note: Prior to the adult stage, metamorphosis occurs during a pupal stage. The adult emerges from the pupal case.

segmented pattern of a *Drosophila* embryo about 10 hours after fertilization. Later in this section, we will explore how the coordination of gene expression underlies the formation of segments.

At the end of **embryogenesis,** which lasts about 18 to 22 hours, a larva hatches from the egg (Figure 24.5f) and begins feeding. *Drosophila* has three larval stages, separated by molts. During molting, the larva sheds its cuticle, a hardened extracellular shell that is secreted by the epidermis.

After the third larval stage, *Drosophila* forms a pupa that undergoes metamorphosis. During metamorphosis, groups of cells called imaginal disks that were produced earlier in development differentiate into the structures found in the adult fly, such as the head, wings, legs, and abdomen. The adult fly emerges from the pupal case about 10 days after fertilization. In *Drosophila*, as in all bilateral animals, the final result of development is an adult body organized along two major axes: the **anteroposterior axis** and the **dorsoventral axis** (Figure 24.5g). The **left-right axis** occurs relative to the anteroposterior axis. An additional axis, the **proximodistal axis,** is used mostly for designating limb parts**.**

Although many interesting developmental events occur during the three larval stages and the pupal stage, we will focus most of our attention on the events that occur during embryonic development. Even before hatching, the embryo develops the basic body plan that is eventually found in the adult organism. In other words, during the early stages of development, the embryo is divided into segments that correspond to the segments of the larva and adult. Therefore, an understanding of how these segments form in the embryo is critical to our understanding of pattern formation.

In *Drosophila*, the establishment of the body axes and division of the body into segments involves the participation of a few dozen genes. **Table 24.1** lists many of the important genes governing pattern formation during embryonic development. These genes were identified by characterizing mutants that had altered development patterns. They are often given interesting names based on the observed phenotype when they are mutant. Rather than examining how all of these genes exert their effects during embryonic development, we will focus on a few examples that illustrate how the expression of a particular gene and the localization of its gene product have a defined effect on the pattern of development.

TABLE **24.1**

Examples of *Drosophila* Genes That Play a Role in Pattern Development

Description	Examples
Maternal effect genes play a role in determining the axes of development. Also certain genes govern the formation of the extreme terminal (anterior and posterior) regions.	Anterior: *bicoid, exuperantia, swallow, staufen* Posterior: *nanos, cappuccino, oskar pumilio, spire, staufen, tudor, vasa* Terminal: *torso, torsolike, trunk, NTF-1* Dorsoventral: *Toll, cactus, dorsal, easter, gurken, nudel, pelle, pipe, snake, spatzle*
Segmentation genes play a role in promoting the subdivision of the embryo into segments. The three types are gap genes, pair-rule genes, and segment-polarity genes.	Gap genes: *empty spiracles, giant, huckebein, hunchback, knirps, Krüppel, tailless, orthodenticle* Pair-rule genes: *even-skipped, hairy, runt, fushi tarazu, paired* Segment-polarity genes: *frizzled, frizzled-2, engrailed, patched, smoothened, hedgehog, wingless, gooseberry*
Homeotic genes play a role in determining the fate of particular segments. *Drosophila* has two clusters of homeotic genes known as the *Antennapedia* complex and the *bithorax* complex.	*Antennapedia* complex: *labial, proboscipedia, Deformed, Sex combs reduced, Antennapedia* *bithorax* complex: *Ultrabithorax, abdominal A, Abdominal B*

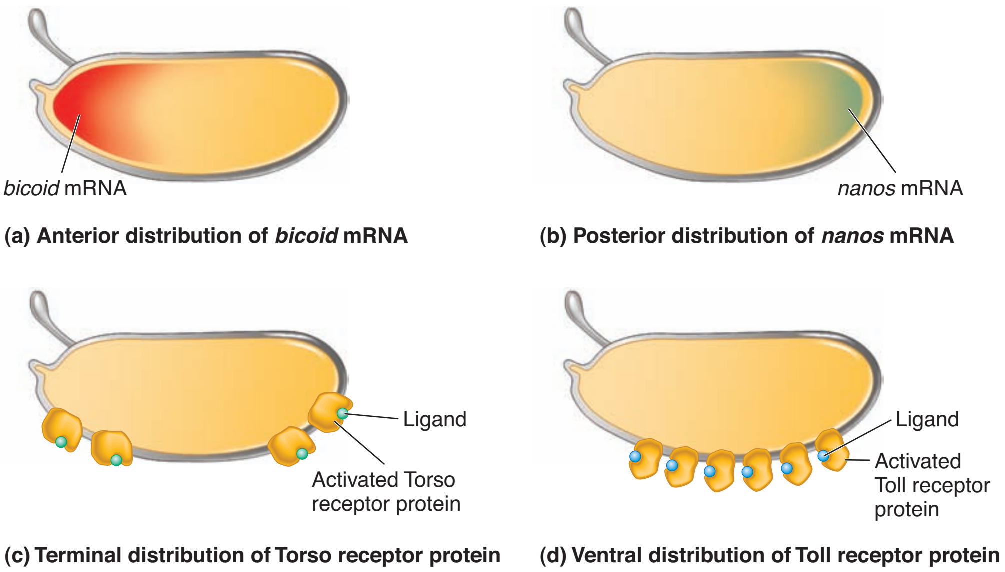
The Gene Products of Maternal Effect Genes Are Deposited Asymmetrically into the Oocyte and Establish the Anteroposterior and Dorsoventral Axes at a Very Early Stage of Development

The first stage in *Drosophila* embryonic pattern development is the establishment of the body axes. This occurs before the embryo becomes segmented. In fact, the morphogens necessary to establish these axes are distributed prior to fertilization. During oogenesis, the mRNA and protein products of genes that are important for the formation of body axes are deposited within the egg. Some of these are deposited asymmetrically, forming morphogen gradients. Later, after the egg has been fertilized and development begins, these gene products establish independent developmental programs that govern the formation of the body axes of the embryo. Proper development depends on the products of many different genes.

As shown in **Figure 24.6**, a few gene products act as key morphogens, or receptors for morphogens, that initiate changes in embryonic development. As shown here, these gene products are deposited asymmetrically in the egg. For example, the product of the *bicoid* gene is necessary to initiate development of the anterior structures of the organism. During oogenesis, the mRNA for *bicoid* accumulates in the anterior region of the oocyte. In contrast, the mRNA from the *nanos* gene accumulates in the posterior end. Later in development, the *nanos* mRNA is translated into protein, which functions to influence posterior development. The *nanos* gene is required for the formation of the abdomen.

In addition to morphogens such as Bicoid and Nanos, the development of the structures at the extreme anterior and posterior ends of the embryo are regulated in part by a receptor protein called Torso. This receptor, which is activated by ligand binding only at the anterior and posterior ends of the egg, is necessary for the formation of the terminal ends of the embryo. The dorsoventral axis is governed by the actions of several proteins, including a receptor protein known as Toll. This receptor is found in plasma membranes throughout the embryo but is activated by ligand binding only along the ventral midline of the embryo.

Let's now take a closer look at the molecular mechanism of one morphogen, Bicoid. The *bicoid* gene got its name because


bicoid mRNA
nanos mRNA
Ligand
Activated Torso receptor protein
Ligand
Activated Toll receptor protein

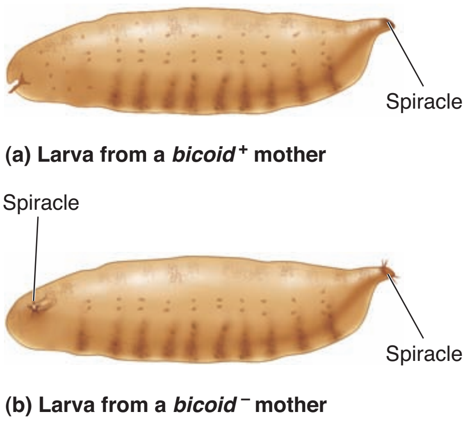
(a) Anterior distribution of *bicoid* mRNA **(b) Posterior distribution of *nanos* mRNA**

(c) Terminal distribution of Torso receptor protein **(d) Ventral distribution of Toll receptor protein**

FIGURE 24.6 **The establishment of the axes of polarity in the *Drosophila* embryo.** This figure shows some of the maternal gene products deposited in the oocyte that are critical in the establishment of the anteroposterior, terminal, and dorsoventral axes. **(a)** *Bicoid* mRNA is distributed in the anterior end of the oocyte and promotes the formation of anterior structures. **(b)** *Nanos* mRNA is localized to the posterior end and promotes the formation of posterior structures. **(c)** The Torso receptor protein is found in the membrane and is activated by ligand binding at either end of the oocyte. It causes the formation of structures that are found only at the ends of the organism. **(d)** The Toll receptor protein is activated by ligand binding at the ventral side of the embryo and establishes the dorsoventral axis. Note: The Torso and Toll receptor proteins are distributed throughout the plasma membranes of the oocyte and embryo, but they are activated by ligand binding only in the regions shown in this figure.

Concept check: Describe the orientations of the anteroposterior and dorsoventral axes.

a larva whose mother is defective in this gene develops with two posterior ends (**Figure 24.7**). This allele exhibits a **maternal effect** pattern of inheritance, in which the genotype of the mother determines the phenotypic traits of the offspring (see Chapter 6). A female fly that is phenotypically normal (because its mother was heterozygous for the normal *bicoid* allele), but genotypically homozygous for an inactive *bicoid* allele (because it inherited an inactive allele from both its mother and father), produces 100% abnormal offspring even when mated to a male that is homozygous for the normal *bicoid* allele.



(a) Larva from a *bicoid*$^+$ mother

(b) Larva from a *bicoid*$^-$ mother

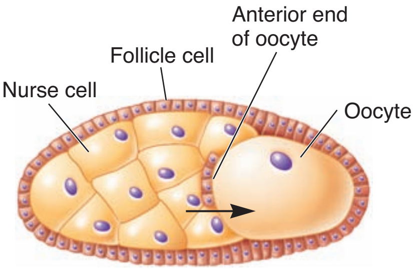
FIGURE 24.7 **The bicoid mutation in *Drosophila*.** **(a)** A larva from a normal *bicoid*$^+$ mother. **(b)** A larva from a homozygous *bicoid*$^-$ mother in which both ends of the larva develop posterior structures. For example, both ends develop a spiracle, which normally is found only at the posterior end. This is a lethal condition.

Concept check: What is the normal function of the bicoid protein?

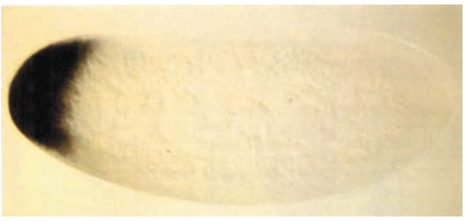
In this example, the *bicoid* mRNA is provided to the oocyte via the nurse cells. In the ovaries of female flies, the nurse cells are localized asymmetrically toward the anterior end of the oocyte. During oogenesis, gene products are transferred from nurse cells into the oocyte via cell-to-cell connections called cytoplasmic bridges. The *bicoid* gene is actively transcribed in the nurse cells, and *bicoid* mRNA is transported into one side of the oocyte, which eventually becomes the anterior end of the embryo (**Figure 24.8a**). The 3′ end of *bicoid* mRNA contains a signal that is recognized by RNA-binding proteins thought necessary for the transport of this mRNA into the oocyte. After it enters the oocyte, the *bicoid* mRNA is trapped at the anterior end.

How can researchers determine the location of *bicoid* mRNA in the oocyte and resulting zygote? **Figure 24.8b** shows the result of an in situ hybridization experiment in which a *Drosophila* egg was examined via a labeled probe complementary to the *bicoid* mRNA. (The technique of in situ hybridization is described in Chapter 21.) As seen here, the *bicoid* mRNA is highly concentrated near the anterior end of the egg cell. When the *bicoid* mRNA is subsequently translated, a gradient of Bicoid protein is established, as shown in **Figure 24.8c**.

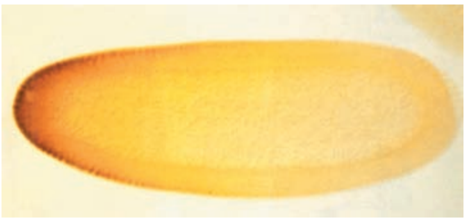
After fertilization occurs, the mRNA is translated into Bicoid protein, which functions as a transcription factor. A remarkable feature of this protein is that its ability to influence gene expression is tuned exquisitely to its concentration. Depending on the concentration of the Bicoid protein, this transcription factor



(a) Transport of maternal effect gene products into the oocyte



(b) In situ hybridization of *bicoid* mRNA



(c) Immunostaining of Bicoid protein

FIGURE 24.8 **Asymmetrical localization of gene products during oogenesis in *Drosophila*. (a)** The nurse cells transport gene products into the anterior (left) end of the developing oocyte. **(b)** An in situ hybridization experiment showing that the *bicoid* mRNA is trapped near the anterior end. **(c)** The *bicoid* mRNA is translated into protein soon after fertilization. The location of the Bicoid protein is revealed by immunostaining using an antibody that specifically recognizes this protein.

Concept check: Where are maternal effect gene products made first? Where do they go?

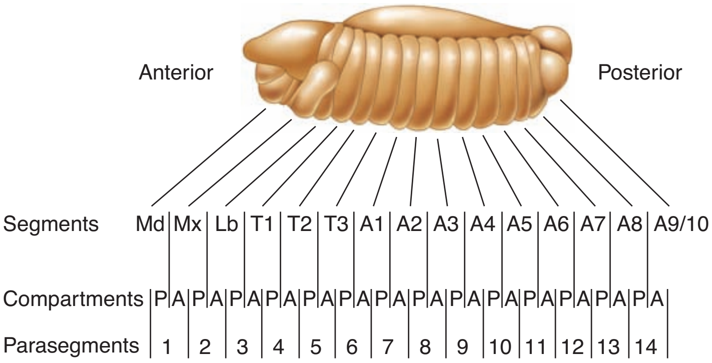


FIGURE 24.9 **A comparison of segments and parasegments in the *Drosophila* embryo.** Note that the parasegments and segments are out of register. The posterior (P) and anterior (A) regions are shown for each segment. The head segments are Md (mandibular), Mx (maxillary) and Lb (labial). The thoracic and abdominal segments are designated with the letters T and A, respectively.

activates genes only in certain regions of the embryo. For example, Bicoid stimulates the transcription of a gene called *hunchback* in the anterior half of the embryo, but its concentration is too low in the posterior half to activate the *hunchback* gene there.

Gap, Pair-Rule, and Segment-Polarity Genes Act Sequentially to Divide the *Drosophila* Embryo into Segments

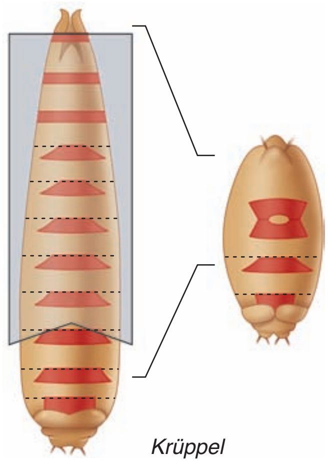
After the anteroposterior and dorsoventral regions of the embryo have been established by maternal effect genes, the next developmental process organizes the embryo into transient subdivisions called parasegments and then permanently into segments. The segmentation pattern of the *Drosophila* embryo is shown in Figure 24.9. This pattern of positional information is maintained, or "remembered," throughout the rest of development. In other words, each segment of the embryo gives rise to unique morphological features in the adult. For example, T2 becomes a thoracic segment with a pair of legs and a pair of wings, and A8 becomes a segment of the abdomen.

Figure 24.9 shows the overlapping relationship between parasegments and segments. As seen here, the boundaries of the segments are out of register with the boundaries of the parasegments. As development proceeds, the anterior compartment of each segment coincides with the posterior region of a parasegment; the posterior compartment of a segment coincides with the anterior region of the next parasegment. The pattern of gene expression that occurs in the posterior region of one parasegment and the anterior region of an adjacent parasegment results in the formation of a segment.

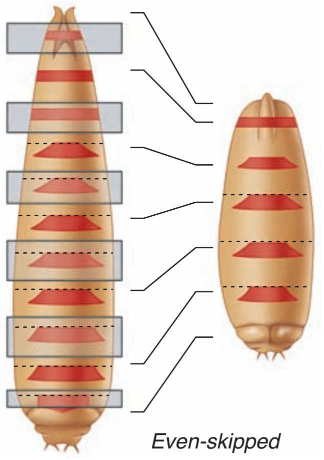
Now that we understand the way the *Drosophila* embryo is subdivided, we can examine how particular genes cause it to become segmented into this pattern. Three classes of genes, collectively called **segmentation genes,** play a role in the formation of body segments: gap genes, pair-rule genes, and segment-polarity genes. The expression and activation patterns of these genes in specific regions of the embryo cause it to become segmented.

How were the three classes of segmentation genes discovered? In the 1970s, segmentation genes were identified by Christiane Nüsslein-Volhard and Eric Wieschaus, who undertook a systematic search for mutations affecting embryonic development. Their pioneering effort identified most of the genes required for the *Drosophila* embryo to develop a segmented pattern.

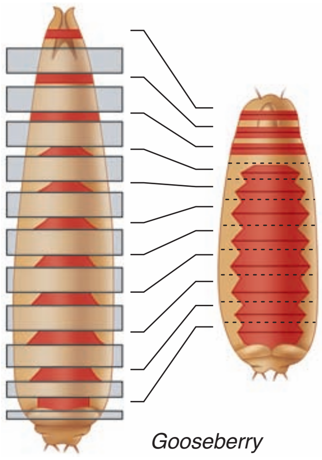
Figure 24.10 presents an example of defects seen in mutants in each of the three types of segmentation genes. The gray boxes indicate the regions missing in the resulting larvae. In flies with a mutation in a gap gene known as *Krüppel*, a contiguous section of the larva is missing (Figure 24.10a). In other words, a gap of several consecutive segments has occurred. By comparison, a defect in a pair-rule gene causes alternating regions to be deleted (Figure 24.10b). For example, when the *even-skipped* gene is defective, portions of alternating segments in the resulting larva are missing. Finally, segment-polarity mutations cause individual segments to be missing either an anterior or a posterior



(a) Gap



(b) Pair-rule



(c) Segment-polarity

FIGURE 24.10 **Phenotypic effects in *Drosophila* larvae with mutations in segmentation genes.** The effects shown here are caused by mutations in **(a)** a gap gene, **(b)** a pair-rule gene, and **(c)** a segment-polarity gene.

Concept check: Describe the different effects of a mutation in a gap gene versus a pair-rule gene.

region. Figure 24.10c shows a mutation in a segmentation gene known as *gooseberry*. When this gene is defective, the anterior portion of each segment is missing from the larva. In the case of segment-polarity mutants, the segments adjacent to the deleted regions exhibit a mirror-image duplication. Overall, the phenotypic effects of mutant segmentation genes provided geneticists with important clues regarding the roles of these genes in the developmental process of segmentation.

Figure 24.11 presents a partial, simplified scheme of the genetic hierarchy that leads to a segmented pattern in the *Drosophila* embryo. Keep in mind that although this figure presents the general sequence of events that occurs during the early stages of embryonic development, many more genes are actually involved in this process (refer back to Table 24.1). As described in this figure, the following steps occur:

1. Maternal effect gene products, such as *bicoid* mRNA, are deposited asymmetrically into the oocyte. These gene products form a gradient that later influences the formation of axes, such as the anteroposterior axis.
2. In contrast to maternal effect genes, which are expressed during oogenesis, **zygotic genes** are expressed after fertilization. Segmentation and homeotic genes are types of zygotic genes. The first zygotic genes to be activated are the gap genes. Maternal effect genes are responsible for activating the gap genes. As shown in step 2 of Figure 24.11, gap genes divide the embryo into a series of broad bands or regions. These bands do not correspond to parasegments or segments within the embryo.
3. The gap genes and maternal effect genes then activate the pair-rule genes. The photograph shown in step 3 of Figure 24.11 illustrates the alternating pattern of expression of the *even-skipped* gene. The expression of pair-rule genes in stripes defines the boundaries of parasegments. *Even-skipped* is expressed in the odd-numbered parasegments. Solved Problem S4 at the end of the chapter examines how the *even-skipped* gene can be expressed in this pattern of alternating stripes. The pair-rule genes divide the broad regions established by gap genes into seven bands or stripes.
4. Once the pair-rule genes are activated in an alternating banding arrangement, their gene products then regulate the expression of segment-polarity genes. The segment-polarity genes divide the embryo into 14 stripes, one within each parasegment. As shown in step 4 of Figure 24.11, the segment-polarity gene *engrailed* is expressed in the anterior region of each parasegment. Another segment-polarity gene, *wingless*, is expressed in the posterior region. Later in development, the anterior end of one parasegment and the posterior end of another parasegment develop into a segment with particular morphological characteristics.

The Expression of Homeotic Genes Controls the Phenotypic Characteristics of Segments

Thus far, we have considered how the *Drosophila* embryo becomes organized along axes and then into a segmented body

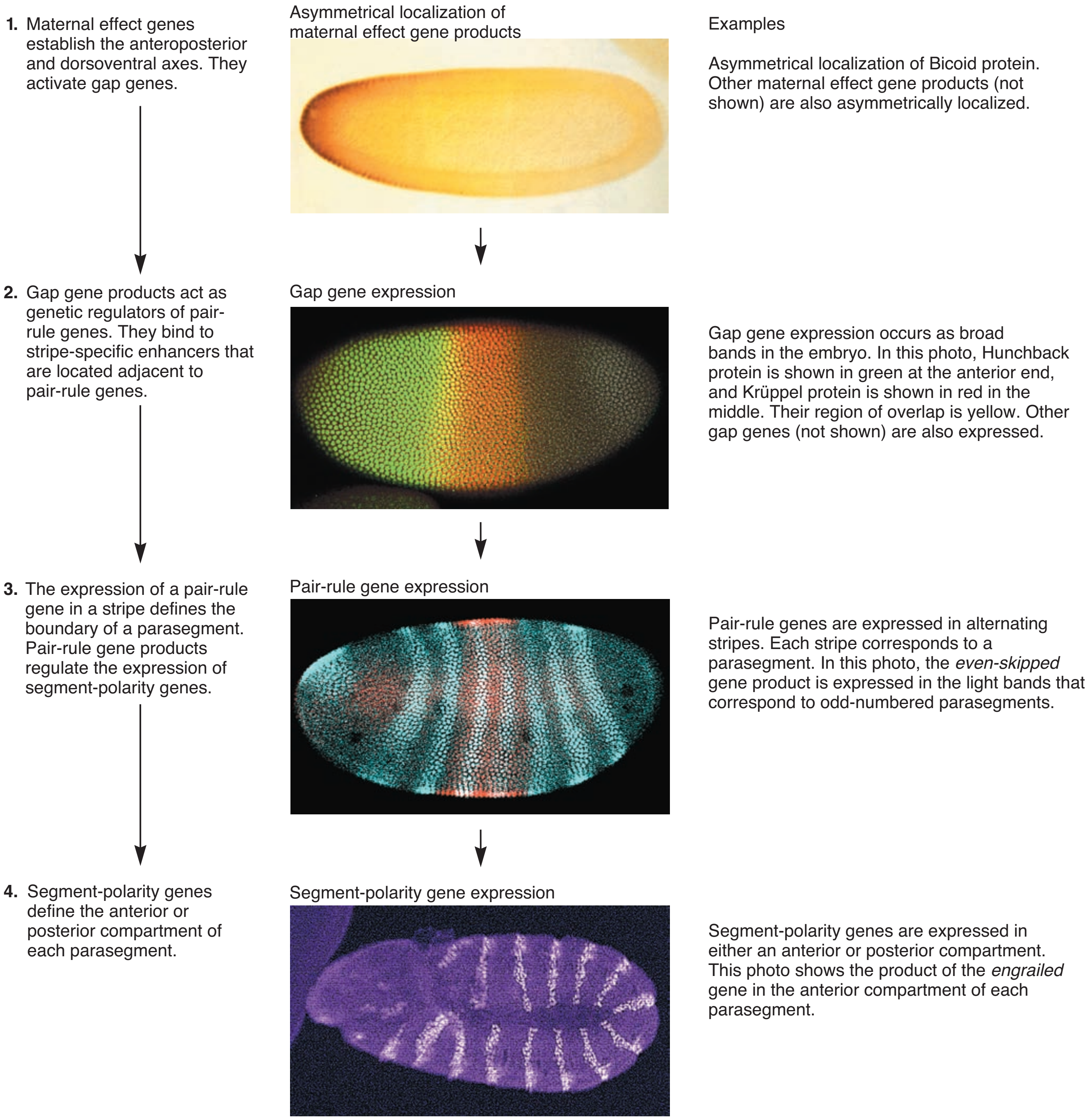

1. Maternal effect genes establish the anteroposterior and dorsoventral axes. They activate gap genes.
Asymmetrical localization of maternal effect gene products
Examples
Asymmetrical localization of Bicoid protein. Other maternal effect gene products (not shown) are also asymmetrically localized.
2. Gap gene products act as genetic regulators of pair-rule genes. They bind to stripe-specific enhancers that are located adjacent to pair-rule genes.
Gap gene expression
Gap gene expression occurs as broad bands in the embryo. In this photo, Hunchback protein is shown in green at the anterior end, and Krüppel protein is shown in red in the middle. Their region of overlap is yellow. Other gap genes (not shown) are also expressed.
3. The expression of a pair-rule gene in a stripe defines the boundary of a parasegment. Pair-rule gene products regulate the expression of segment-polarity genes.
Pair-rule gene expression
Pair-rule genes are expressed in alternating stripes. Each stripe corresponds to a parasegment. In this photo, the *even-skipped* gene product is expressed in the light bands that correspond to odd-numbered parasegments.
4. Segment-polarity genes define the anterior or posterior compartment of each parasegment.
Segment-polarity gene expression
Segment-polarity genes are expressed in either an anterior or posterior compartment. This photo shows the product of the *engrailed* gene in the anterior compartment of each parasegment.

FIGURE 24.11 **Overview of the genetic hierarchy leading to segmentation in *Drosophila*.** Note: When comparing steps 3 and 4, the elongated embryo bends in the middle and folds back on itself.

pattern. Now let's examine how each segment develops its unique morphological features. Geneticists often use the term **cell fate** to describe the ultimate morphological features that a cell or group of cells adopt. For example, the fate of the cells in segment T2 in the *Drosophila* embryo is to develop into a thoracic segment containing two legs and two wings. In *Drosophila*, the fate of the cells in each segment of the body is determined at a very early stage of embryonic development, long before the morphological features become apparent.

Our understanding of developmental fate has been greatly aided by the identification of mutant genes that alter cell fates. In animals, the first mutant of this type was described by the German entomologist Ernst Gustav Kraatz in 1876. He observed a sawfly (*Cimbex axillaris*) in which part of an antenna was replaced with a leg. During the late nineteenth century, the English zoologist William Bateson collected many of these types of observations and published them in 1894 in a book entitled *Materials for the Study of Variation Treated with Especial Regard to Discontinuity in the*

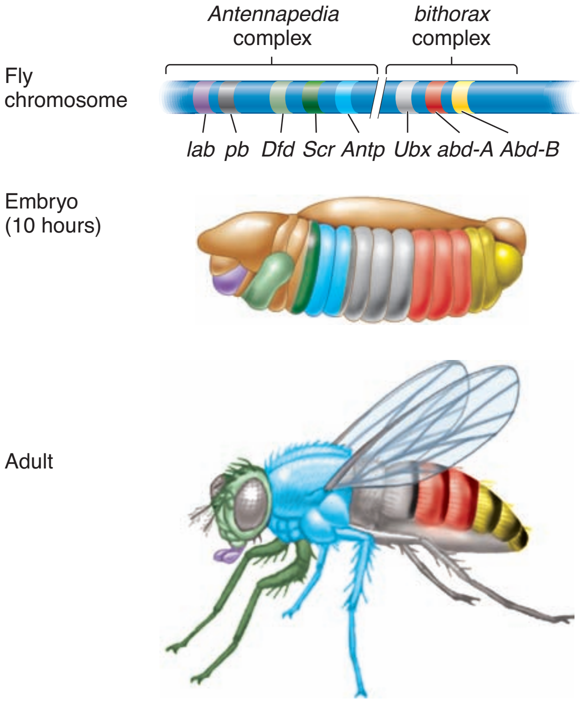
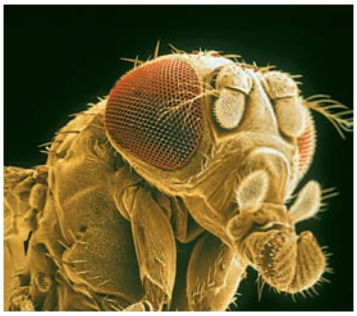
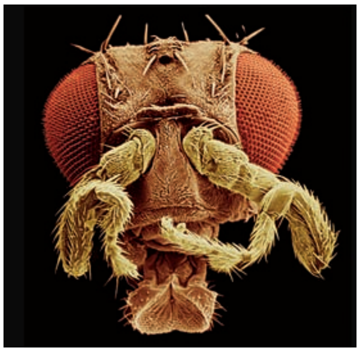


FIGURE 24.12 Expression pattern of homeotic genes in *Drosophila*. The order of homeotic genes, *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal A* (*abd-A*), and *Abdominal B* (*Abd-B*), correlates with the order of expression in the embryo. The expression pattern of seven of these genes is shown. *Lab* (purple) is expressed in the region that eventually gives rise to mouth structures. *Dfd* (light green) is expressed in the region that will form much of the head. *Scr* (forest green) and *Antp* (blue) are expressed in embryonic segments that give rise to thoracic segments. *Ubx* (gray), *abd-A* (red), and *Abd*-B (yellow) are expressed in posterior segments that will form the abdomen. The order of gene expression, from anterior to posterior, parallels the order of genes along the chromosome. The expression pattern of the *pb* gene is not well understood.

Concept check: Explain how the physical arrangements of the homeotic genes correlate with their effects on phenotype.



(a) Normal fly



(b) *Antennapedia* mutant

FIGURE 24.13 The *Antennapedia* mutation in *Drosophila*.

Genes → Traits **(a)** A normal fly with antennae. **(b)** This mutant fly has a gain-of-function mutation in which the *Antp* gene is expressed in the embryonic segment that normally gives rise to antennae. The expression of *Antp* causes this region to have legs rather than antennae.

Origin of Species. In this book, Bateson coined the term **homeotic** to describe mutants in which one body part is replaced by another, such as the transformation of the antennae of insects into legs.

As mentioned earlier, Edward Lewis began to study strains of *Drosophila* having homeotic mutations. This work, which began in 1946, was the first systematic study of homeotic genes. Each homeotic gene controls the fate of a particular region of the body. *Drosophila* contains two clusters of homeotic genes called the *Antennapedia* complex and the *bithorax* complex. Figure 24.12 shows the organization of genes within these complexes. As discussed later in this chapter, these types of genes are found in all animals and referred to as ***Hox* genes.** The *Antennapedia* complex contains five genes, designated *lab*, *pb*, *Dfd*, *Scr*, and *Antp*. The *bithorax* complex has three genes, *Ubx*, *abd-A*, and *Abd-B*. Both the *Antennapedia* and *bithorax* complexes are located on chromosome 3, but a large segment of DNA separates them.

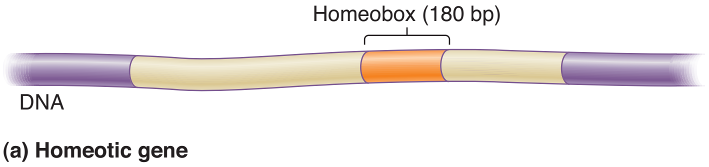
As noted in Figure 24.12, the order of these genes along chromosome 3 correlates with their pattern of gene expression along the anteroposterior axis of the body. This phenomenon is called colinearity. For example, *lab* is expressed in an anterior segment and governs the formation of mouth structures. The *Antp* gene is expressed strongly in the thoracic region during embryonic development and controls the formation of thoracic structures. Transcription of the *Abd-B* gene occurs in the posterior region of the embryo. This gene controls the formation of the posterior-most abdominal segments.

The role of homeotic genes in determining the identity of particular segments has been revealed by mutations that alter their function. For example, the *Antp* gene is normally expressed in the thoracic region. The *Antennapedia* mutation causes the *Antp* gene to also be expressed in the region where the antennae are made. These mutant flies have legs in the place of antennae (Figure 24.13)! This is an example of a **gain-of-function mutation.** In this case, the *Antp* gene is also expressed abnormally in the anterior segment that normally gives rise to the antennae.

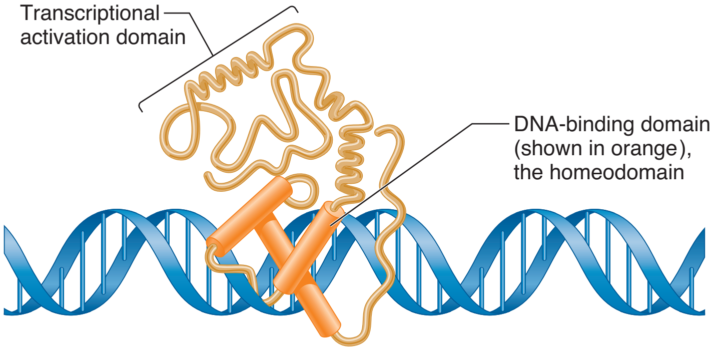
Investigators have also studied many loss-of-function mutations in homeotic genes. When a particular homeotic gene is defective, the region it normally governs is usually controlled by the homeotic gene that acts in the adjacent anterior region. For example, the *Ubx* gene functions within parasegments 5 and 6. If this gene is missing, this section of the fly becomes converted to the structures found in parasegment 4.

The homeotic genes are part of the genetic hierarchy that produces the morphological characteristics of the fly. How are they regulated? The homeotic genes are controlled by gap genes and pair-rule genes, and they are also regulated by interactions among themselves. Other genes maintain the patterns of homeotic gene expression after the initial patterns are established by segmentation genes. A group of genes known as *Polycomb* genes represses the expression of homeotic genes in regions of the embryo where they should not act. This is accomplished by the remodeling of chromatin structure to convert it to a closed conformation in which transcription is difficult or impossible. Alternatively, in regions where the homeotic genes are active, the products of *Trithorax* genes maintain an open chromatin conformation in which transcription can take place. Overall, the concerted actions of many gene products cause the homeotic genes to be expressed only in the appropriate region of the embryo, as shown in Figure 24.12.

Because they are part of a genetic hierarchy in which genes activate other genes, perhaps it is not surprising that homeotic genes encode transcription factors. The coding sequence of homeotic genes contains a 180-bp consensus sequence known as a **homeobox** (Figure 24.14a). This sequence was first discovered in the *Antp* and *Ubx* genes, and it has since been found in all *Drosophila* homeotic genes and in some other genes affecting pattern development, such as *bicoid*. The protein domain encoded by the homeobox is called a **homeodomain.** The arrangement of α helices within the homeodomain promotes the binding of the protein to the major groove of DNA (Figure 24.14b). In this way, homeotic proteins can bind to DNA in a sequence-specific manner. In addition to DNA-binding ability, homeotic proteins also contain a transcriptional activation domain that functions to activate the genes to which the homeodomain can bind.



(a) Homeotic gene


Transcriptional activation domain
DNA-binding domain (shown in orange), the homeodomain

(b) Homeotic protein bound to DNA

FIGURE 24.14 Molecular features of homeotic proteins. **(a)** A homeotic gene (shown in tan and orange) contains a 180-bp sequence called the homeobox (shown in orange). **(b)** When a homeotic gene is expressed, it produces a protein that functions as a transcription factor. The homeobox encodes a region of the protein called a homeodomain, which binds to the major groove of DNA. These DNA-binding sites are found within genetic regulatory elements (i.e., enhancers). The enhancers are found in the vicinity of promoters that are turned on by homeotic proteins. For this to occur, the homeotic protein also contains a transcriptional activation domain, which activates the transcription of a gene after the homeodomain has bound to the DNA.

The transcription factors encoded by homeotic genes activate genes encoding proteins that produce the morphological characteristics of each segment. Much current research attempts to identify these genes and determine how their expression in particular regions of the embryo leads to morphological changes in the embryo, larva, and adult. In some cases, these genes also encode transcription factors that control the expression of other sets of genes that alter the morphological characteristics of cells. In other cases, genes that are activated by homeotic proteins play a role in cell-to-cell signaling pathways. The activation of these signaling pathways enables individual cells and groups of cells to adopt their correct morphologies.

The Developmental Fate of Each Cell in *Caenorhabditis elegans* Is Known

We now turn our attention to another invertebrate, *C. elegans*, a nematode that has been the subject of numerous studies in developmental genetics. The embryo develops within the eggshell and hatches when it reaches a size of 550 cells. After hatching, it continues to grow and mature as it passes through four successive molts. It takes about 3 days for a fertilized egg to develop into an adult worm about 1 mm in length. With regard to sex, *C. elegans* can be a male (and produce only sperm) or a hermaphrodite (capable of producing sperm and egg cells). An adult male is composed of 1031 somatic cells and produces about 1000 sperm. A hermaphrodite consists of 959 somatic cells and produces about 2000 gametes (both sperm and eggs).

A remarkable feature of this organism is that the pattern of cellular development is extremely constant from worm to worm. In the early 1960s, Sydney Brenner pioneered the effort to study the pattern of cell division in *C. elegans* and establish it as a model organism. Because *C. elegans* is transparent and composed of relatively few cells, researchers can follow cell division step by step under the microscope, beginning with a fertilized egg and ending with an adult worm. Researchers can identify a particular cell at an embryonic stage, follow that cell as it divides, and observe where its descendant cells are located in the adult. An illustration that depicts how cell division proceeds is called a **cell lineage diagram.** It describes the cell division patterns and fates of any cell's descendants.

Figure 24.15 shows a partial cell lineage diagram for a *C. elegans* hermaphrodite. At the first cell division, the egg divides to produce two cells, called AB and P_1. AB then divides into two cells—ABa and ABp; and P_1 divides into two cells—EMS and P_2. The EMS cell then divides into two cells, called MS and E. The cellular descendants of the E cell give rise to the worm's intestine. In other words, the fate of the E cell's descendants is to develop

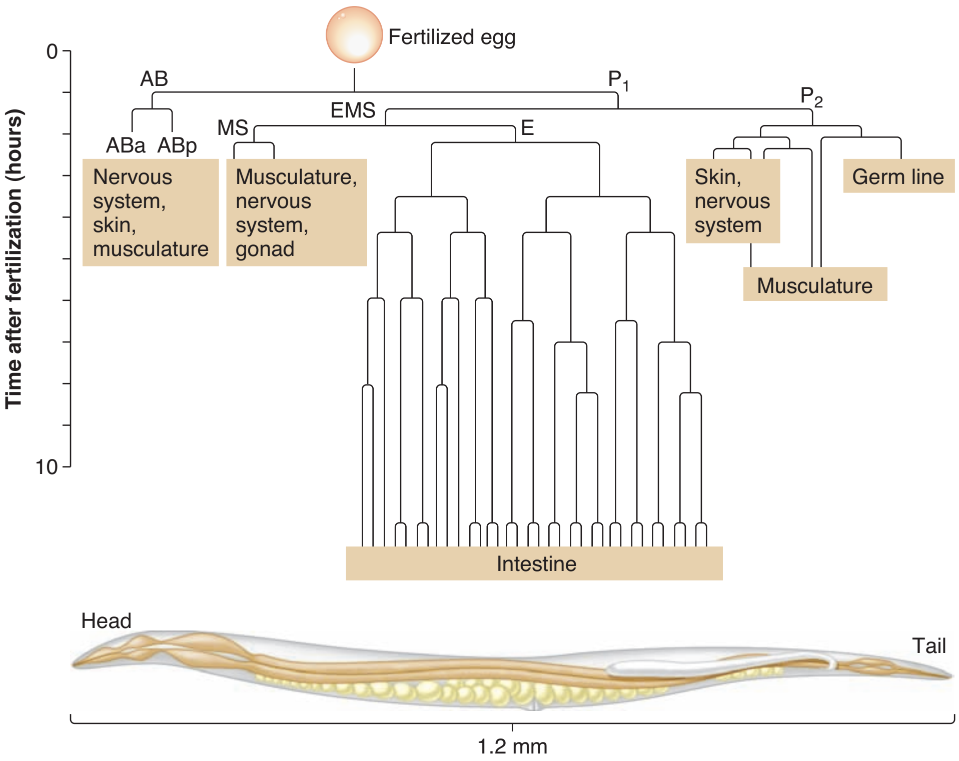


FIGURE 24.15 **A cell lineage diagram of the nematode *Caenorhabditis elegans*.** This partial cell lineage diagram illustrates how the cells divide to produce different regions of the adult. The fate of the intestinal cell lineage is shown in greater detail than that of other cell lineages. A complete lineage diagram is known for this organism, although its level of detail is beyond the scope of this textbook. Note: The lowercase letters "a" and "p" refer to anterior and posterior, respectively, and thereby indicate the relative positions of the daughter cells.

Concept check: *What is a cell lineage?*

into intestinal cells. This diagram also illustrates the concept of a **cell lineage**—a series of cells that are derived from each other by cell division. For example, the EMS cell, E cell, and the intestinal cells are all part of the same cell lineage. Why is a cell lineage diagram for an organism an important experimental advantage? The entire genome sequence of *C. elegans* is known, and researchers have identified many genes that affect development. By analyzing mutants in a particular gene, a cell lineage diagram enables researchers to investigate how gene expression in any cell, at any stage of development, is important for the outcome of a cell's fate.

24.2 REVIEWING THE KEY CONCEPTS

- *Drosophila* proceeds through different developmental stages from a fertilized egg to an adult. Various sets of genes are responsible for developmental changes (see Figure 24.5, Table 24.1).
- Maternal effect genes establish the anteroposterior and dorsoventral axes due to their asymmetrical distribution. An example is *bicoid*, which promotes the formation of anterior structures (see Figures 24.6–24.8).
- The *Drosophila* embryo is divided into segments (see Figure 24.9).
- Three categories of segmentation genes have been identified based on analyzing the phenotypes of mutations affecting segmentation (see Figure 24.10).
- A hierarchy of gene expression, which includes maternal effect genes, gap genes, pair-rule genes, and segment-polarity genes, gives rise to a segmented *Drosophila* embryo (see Figure 24.11).
- Homeotic genes control the developmental fate of particular segments (see Figures 24.12, 24.13).
- Homeotic proteins contain a DNA-binding domain and a transcriptional activation domain (see Figure 24.14).
- In *C. elegans*, a cell lineage diagram describes the cell fate of each cell of the worm's body (see Figure 24.15).

24.2 COMPREHENSION QUESTIONS

1. The expression of maternal effect genes directly leads to
    a. the establishment of body axes.
    b. segmentation.
    c. determination.
    d. cell differentiation.
2. Which of the following are types of segmentation genes?
    a. Gap genes
    b. Pair-rule genes
    c. Segment-polarity genes
    d. All of the above
3. The expression of homeotic genes leads to
    a. the establishment of body axes.
    b. the formation of segments in the embryo.
    c. the determination of structures within segments.
    d. cell differentiation.
4. Homeotic genes encode proteins that function as
    a. cell-signaling proteins.
    b. transcription factors.
    c. hormones.
    d. all of the above.

24.3 VERTEBRATE DEVELOPMENT

Biologists have studied the morphological features of development in many vertebrate species. Historically, amphibians and birds have been studied extensively, because their eggs are rather large and easy to manipulate. For example, the developmental stages of the frog and chicken have been described in great detail. In more recent times, the successes obtained in *Drosophila* have shown the great power of genetic analyses in elucidating the underlying molecular mechanisms that control biological development. With this knowledge, many researchers are attempting to understand the genetic pathways that govern the development of the more complex body structure found in vertebrate organisms.

Several vertebrate species have been the subject of developmental studies. These include the mouse (*Mus musculus*), the frog (*Xenopus laevis*), and the small aquarium zebrafish (*Danio rerio*). In this section, we will primarily discuss the genes that are important in mammalian development, particularly those that have been characterized in the mouse, one of the best-studied mammals. As we will see, several genes affecting its developmental pathways have been cloned and characterized. In this section, we will examine how these genes affect the course of vertebrate development.

Researchers Have Identified Homeotic Genes in Vertebrates

When two or more genes have similar sequences because they are derived from the same ancestral gene, they are called **homologous genes.** Homologous genes found in different species are termed **orthologs.** Researchers have found complexes of homeotic genes in many vertebrate species that are homologous to those in the fruit fly. These groups of adjacent homeotic genes are called ***Hox* complexes.** Like their counterparts in *Drosophila*, the *Hox* genes encode transcription factors that regulate many different genes. As shown in **Figure 24.16**, the mouse has four *Hox* complexes, designated *HoxA* (on chromosome 6), *HoxB* (on chromosome 11), *HoxC* (on chromosome 15), and *HoxD* (on chromosome 2). Thirteen different types of homeotic genes occur within the four *Hox* complexes, although none of the four complexes contains representatives of all 13 types of genes. A total of 38 genes are found in the four complexes. The addition of *Hox* genes into the genomes of animals has allowed animals to develop more complex body plans. We will consider the evolutionary origin of the *Hox* genes and their importance in the evolution of animal species in Chapter 27 (see Figure 27.17).

Remarkably, several of the homeotic genes in fruit flies and mammals are strikingly similar. Among the first six types of genes, five of them are homologous to genes found in the *Antennapedia* complex of *Drosophila* (1, 2, 4, 5, and 6). Among the last seven, three are homologous to the genes of the *bithorax* complex (7, 8, and 9). These results indicate that fundamental similarities occur in the ways that animals as different as fruit flies and mammals undergo embryonic development. This suggests that a universal body plan underlies animal development.

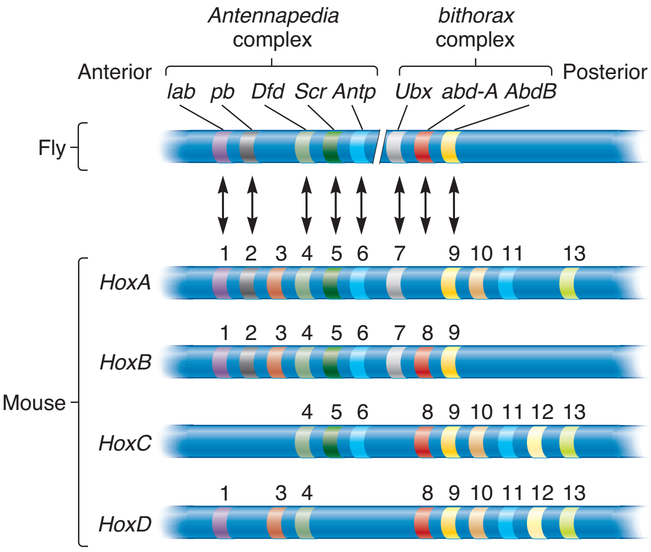


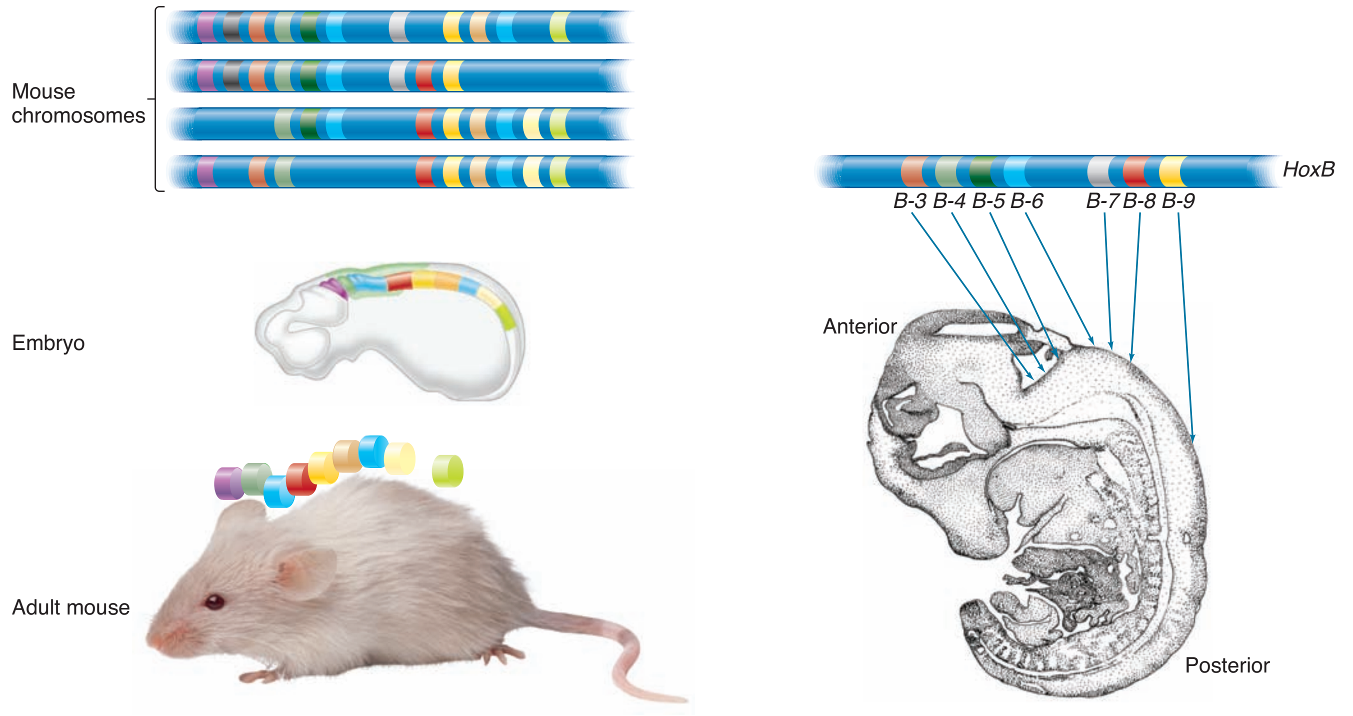
FIGURE 24.16 **A comparison of homeotic genes in *Drosophila* and the mouse.** The mouse contains four gene clusters, *HoxA–D*, that correspond to certain homeotic genes found in *Drosophila*. Thirteen different types of homeotic genes are found in the mouse, although each *Hox* gene cluster does not contain all 13 genes. In this drawing, orthologous genes are aligned in columns. For example, *lab* is the ortholog to *HoxA-1*, *HoxB-1*, and *HoxD-1*; *Ubx* is orthologous to *HoxA-7* and *HoxB-7*.

Concept check: *What is an ortholog?*

Like the *Antennapedia* and *bithorax* complexes in *Drosophila*, the arrangement of *Hox* genes along the mouse chromosomes reflects their pattern of expression from the anterior to the posterior end (**Figure 24.17a**). This phenomenon is seen in more detail in **Figure 24.17b**, which shows the expression pattern for a group of *HoxB* genes in a mouse embryo. Overall, these results are consistent with the idea that the *Hox* genes play a role in determining the fates of segments along the anteroposterior axis.

Currently, researchers are trying to understand the functional roles of the genes within the *Hox* complexes in vertebrate development. In *Drosophila*, great advances in developmental genetics have been made by studying mutant alleles in genes that control development. In mice, however, few natural mutations have been identified that affect development. This has made it difficult to understand the role that genetics plays in the development of the mouse and other vertebrate organisms.

How have researchers overcome this problem? Geneticists are taking an approach known as **reverse genetics.** In this strategy, researchers first identify the wild-type gene using cloning methods. In this case, the *Hox* genes in vertebrates were initially cloned using *Drosophila* genes as probes. The next step is to create a mutant version of a *Hox* gene in vitro. This mutant allele is then reintroduced into a mouse using gene replacement techniques described in Chapter 20. When the function of a wild-type gene is eliminated, it is called a **gene knockout.** In this



(a) Correlation between *Hox* gene arrangement and expression

(b) Anterior expression boundaries for a series of *HoxB* genes

FIGURE 24.17 **Expression pattern of *Hox* genes in the mouse.** **(a)** A schematic illustration of the *Hox* gene expression in the embryo and the corresponding regions in the adult. **(b)** A more detailed description of *HoxB* expression in a mouse embryo. The arrows indicate the anterior-most boundaries for the expression of *HoxB-3* to *HoxB-9*. The order of *Hox* gene expression, from the anterior to posterior end of the embryo, is in the same order as the genes are found along the chromosome.

way, researchers can determine how the mutant allele affects the phenotype of the mouse.

The term *reverse genetics* reflects the idea that the experimental steps occur in an order opposite to that in the conventional approach used in *Drosophila* and other organisms. In the fly, mutant alleles were identified by their phenotype first, and then they were cloned. In the mouse, the genes were cloned first, the mutations were made in vitro, and then the mutated genes were introduced into the mouse to observe their phenotypic effects.

Overall, the current picture indicates that the *Hox* genes in vertebrates play a key role in determining the fates of structures found along the anteroposterior axis. During the evolution of animals, increases in the number of *Hox* genes and changes in their patterns of expression have had an important effect on their morphologies. As an example, let's consider the expression of the *HoxC-6* gene. Changes in its pattern of expression among vertebrate species are associated with changes in the boundary between cervical (neck) vertebrae and thoracic (chest) vertebrae. The *HoxC-6* gene is expressed during embryonic development prior to vertebrae formation. Differences in the relative position of its expression correlate with the number of neck vertebrae produced (**Figure 24.18**). In the mouse, which has a relatively short neck, *HoxC-6* expression begins in the region of the early embryo that later develops into vertebrae 7 and 8. In contrast, *HoxC-6* expression in the chicken and goose begins much farther back, between vertebrae 14 and 15, or 17 and 18, respectively. The forelimbs also arise at this boundary in all vertebrates. However, snakes, which have no neck or forelimbs, do not have such a boundary because *HoxC-6* expression begins toward their heads.

Genes That Encode Transcription Factors Also Play a Key Role in Cell Differentiation

Thus far, we have focused our attention on patterns of gene expression that occur during the very early stages of development. These genes control the basic body plan of the organism. As this process occurs, cells become **determined.** As mentioned earlier, this refers to the phenomenon that a cell is destined to become a particular cell type. In other words, its fate has been predetermined to eventually become a particular type of cell such as a nerve cell. This occurs long before a cell becomes **differentiated.** This later term means that a cell's morphology and function have changed, usually permanently, into a highly specialized cell type. For example, an undifferentiated mesodermal cell may differentiate into a specialized muscle cell, or an ectodermal cell may differentiate into a nerve cell.

At the molecular level, the profound morphological differences between muscle cells and nerve cells arise from gene regulation. Though muscle and nerve cells contain the same set of genes, the cells regulate the expression of those genes in very different ways. Certain genes that are transcriptionally active in

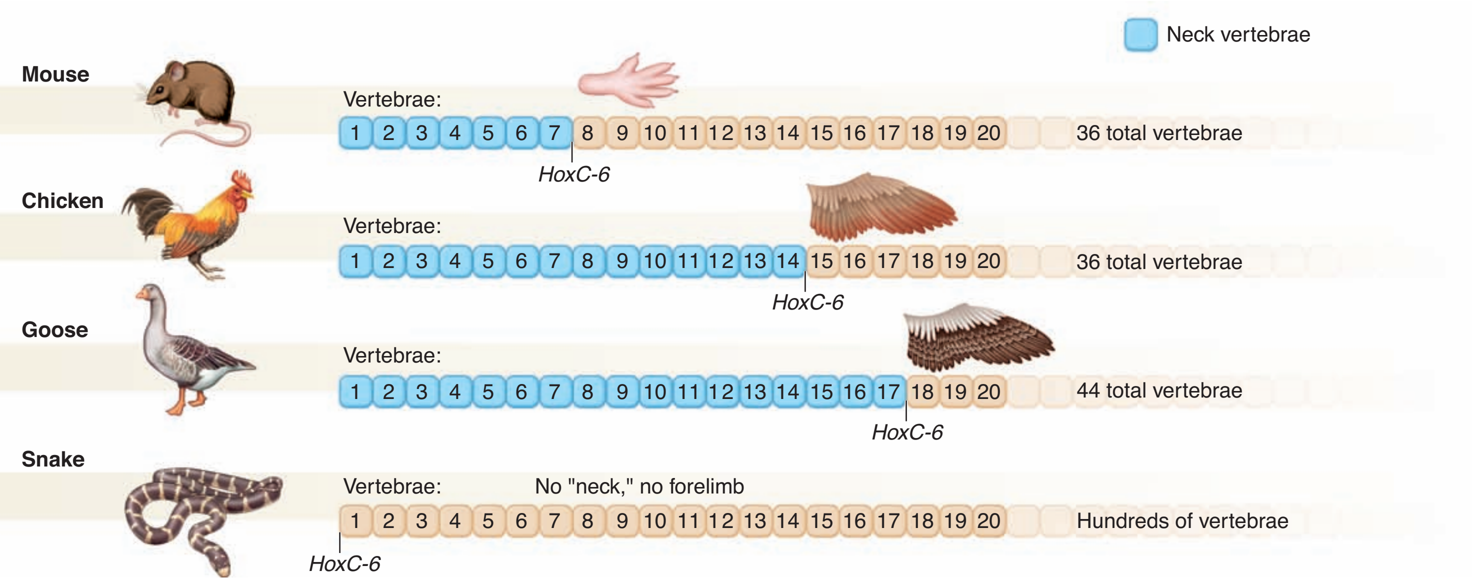


FIGURE 24.18 **Expression of the *HoxC-6* gene in different species of vertebrates.**
This figure indicates the anterior boundary where the *HoxC-6* gene is expressed in the mouse, chicken, goose, and snake.

Concept check: *Briefly describe how the HoxC-6 gene affects vertebrate development.*

muscle cells are inactive in nerve cells, and vice versa. Therefore, muscle and nerve cells express different proteins, which affect the morphological and physiological characteristics of the respective cells in distinct ways. In this manner, differential gene regulation underlies cell differentiation.

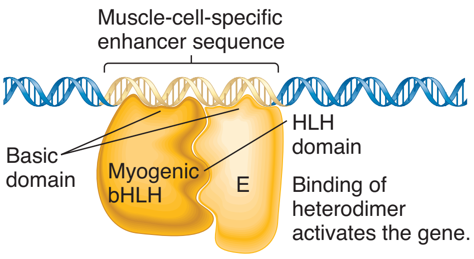
We learned earlier that a hierarchy of gene regulation is responsible for establishing the body pattern in *Drosophila*. Maternal effect genes control the expression of gap genes, which control the expression of pair-rule genes, and so forth. A similar type of hierarchy is thought to underlie cell differentiation. Researchers have identified specific genes that cause cells to differentiate into particular cell types. These genes trigger undifferentiated cells to differentiate and follow their proper cell fates.

In 1987, Harold Weintraub and colleagues identified a gene, which they called *MyoD*. This gene plays a key role in skeletal muscle cell differentiation. Experimentally, when the cloned *MyoD* gene was introduced into fibroblast cells in a laboratory, the fibroblasts differentiated into skeletal muscle cells. This result was particularly remarkable because fibroblasts normally differentiate into osteoblasts (bone cells), chondrocytes (cartilage cells), adipocytes (fat cells), and smooth muscle cells, but in vivo, they never differentiate into skeletal muscle cells.

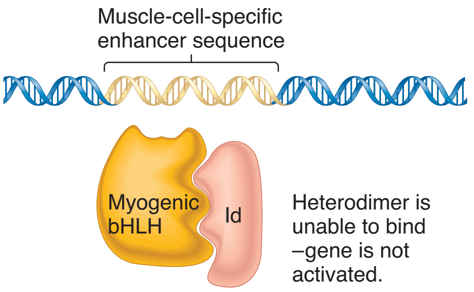
Since this initial discovery, researchers have found that *MyoD* belongs to a small group of genes that initiate muscle development. Besides *MyoD*, these include *Myogenin*, *Myf5*, and *Mrf4*. All four of these genes encode transcription factors that contain a **basic domain** and a **helix-loop-helix domain (bHLH).** The basic domain is responsible for DNA binding and the activation of skeletal muscle-cell-specific genes. The helix-loop-helix domain is necessary for dimer formation between transcription factor proteins. Because of their common structural features and their role in muscle differentiation, MyoD, Myogenin, Myf5, and Mrf4 constitute a family of proteins called **myogenic bHLH proteins.** They are found in all vertebrates and have been identified in several invertebrates, such as *Drosophila* and *C. elegans*. In all cases, the myogenic bHLH genes are activated during skeletal muscle cell development.

At the molecular level, certain key features enable myogenic bHLH proteins to promote muscle cell differentiation. The basic domain binds to a muscle-cell-specific enhancer sequence; this sequence is adjacent to genes that are expressed only in muscle cells (**Figure 24.19**). Therefore, when myogenic bHLH proteins are activated, they can bind to these enhancers and activate the expression of many different muscle-cell-specific genes. They may exert their effects via alterations in chromatin structure or via the activation of RNA polymerase to a transcriptionally active state. In this way, myogenic bHLH proteins function as master switches that activate the expression of many muscle-specific genes. When the encoded proteins are synthesized, they change the characteristics of an undifferentiated cell into those of a highly specialized skeletal muscle cell.

Another important aspect of myogenic bHLH proteins is that their activity is regulated by dimerization, which occurs via the helix-loop-helix domain. As shown in Figure 24.19, heterodimers—dimers formed from two different proteins—may be activating or inhibitory. When a heterodimer forms between a myogenic bHLH protein and an E protein, which also contains a basic domain, the heterodimer binds to the DNA and activates gene expression (Figure 24.19a). However, when a heterodimer forms between a myogenic bHLH protein and a protein called Id (for inhibitor of differentiation), the heterodimer cannot bind to DNA because the Id protein lacks a basic domain (Figure 24.19b). Two basic domains are needed for the heterodimer to bind to the DNA. The Id protein is produced during early


Muscle-cell-specific enhancer sequence
Basic domain
Myogenic bHLH
E
HLH domain
Binding of heterodimer activates the gene.

(a) Action of bHLH–E heterodimer


Muscle-cell-specific enhancer sequence
Myogenic bHLH
Id
Heterodimer is unable to bind –gene is not activated.

(b) Action of bHLH–Id heterodimer

FIGURE 24.19 **Regulation of muscle-cell-specific genes by myogenic bHLH proteins. (a)** A heterodimer formed from a myogenic bHLH protein and an E protein can bind to a muscle-cell-specific enhancer sequence and activate gene expression. **(b)** A heterodimer formed from a myogenic bHLH protein and an Id protein cannot bind to the DNA and therefore does not activate gene transcription.

Concept check: At which stage of development does the Id protein function? Why is its function important?

stages of development and prevents myogenic bHLH proteins from promoting muscle differentiation too soon. At later stages of development, the amount of Id protein falls, and myogenic bHLH proteins can then combine with E proteins to induce muscle-cell differentiation.

24.3 REVIEWING THE KEY CONCEPTS

- Homeotic genes in vertebrates are found in *Hox* complexes (see Figure 24.16).
- *Hox* genes control the fate of regions along the anteroposterior axis (see Figures 24.17, 24.18).
- Transcription factors also control cell differentiation. An example is *MyoD*, which causes cells to differentiate into skeletal muscle cells (see Figure 24.19).

24.3 COMPREHENSION QUESTIONS

1. *Hox* genes encode transcription factors that
   a. control segmentation.
   b. promote determination.
   c. cause cell differentiation.
   d. do all of the above.
2. A cell that is ___________ has a particular morphology and function.
   a. determined
   b. differentiated
   c. undergoing apoptosis
   d. dividing
3. Myogenic bHLH proteins are ___________ that promote ___________.
   a. cell-signaling proteins, muscle cell differentiation
   b. cell-signaling proteins, muscle cell determination
   c. transcription factors, muscle cell differentiation
   d. transcription factors, muscle cell determination

24.4 PLANT DEVELOPMENT

In developmental plant biology, the model organism for genetic analysis is *Arabidopsis thaliana* (**Figure 24.20**). Unlike most flowering plants, which have long generation times and large genomes, *Arabidopsis* has a generation time of about 2 months and a genome size of 14 × 10^7 bp, which is similar to the genomes of *Drosophila* and *C. elegans*. A flowering *Arabidopsis* plant produces a large number of seeds and is small enough to be grown in the laboratory. Like *Drosophila*, *Arabidopsis* can be subjected to mutagens to generate mutations that alter developmental processes. The small genome size of this organism makes it relatively easy to map these mutant alleles and eventually clone the relevant genes.

The morphological patterns of growth are markedly different between animals and plants. As described previously, animal embryos become organized along anteroposterior, dorsoventral, and left-right axes, and then they subdivide into segments. By comparison, the form of plants has two key features. The first is the root-shoot axis. Most plant growth occurs via cell division near the tips of the shoots and the bottoms of the roots.

Second, this growth occurs in a well-defined radial pattern. For example, early in *Arabidopsis* growth, a rosette of leaves or flowers is produced from buds that emanate in a spiral pattern directly from the main shoot (see Figure 24.20). Later, the shoot generates branches that also produce leaf buds as they grow. Overall, the radial pattern in which a plant shoot gives off the buds that produce leaves, flowers, and branches is an important mechanism that determines much of the general morphology of the plant.

At the cellular level, too, plant development differs markedly from animal development. For example, cell migration does not occur during plant development. In addition, the development of a plant does not rely on morphogens that are deposited asymmetrically in the oocyte. In plants, an entirely new individual can be regenerated from many types of somatic cells. In other words, many plant cells are **totipotent,** meaning that they have the ability to differentiate into every cell type and to produce an entire individual. By comparison, animal development typically relies on the organization within an oocyte as a starting point for development.

In spite of these apparent differences, the underlying molecular mechanisms of pattern development in plants still share some similarities with those in animals. In this section, we will consider a few examples in which genes encoding transcription factors play a key role in plant development.

FIGURE 24.20 **The model organism *Arabidopsis*.** The plant is relatively small, making it easy to grow many of them in the laboratory.

Plant Growth Occurs from Meristems Formed During Embryonic Development

Figure 24.21 illustrates a common sequence of events that takes place in the development of seed plants such as *Arabidopsis*. After fertilization, the first cellular division is asymmetrical and produces a smaller cell, called the apical cell, and a larger basal cell (Figure 24.21a). The apical cell gives rise to most of the embryo, and it later develops into the shoot of the plant. The basal cell gives rise to the root, along with the suspensor that produces extraembryonic tissue required for seed formation. At the "heart" stage, which is composed of only about 100 cells, the basic organization of the plant has been established (Figure 24.21d). The **shoot meristem** arises from a group of cells located between the cotyledons, which are embryonic leaves. These cells are the precursors that will produce the shoot of the plant, along with lateral structures such as leaves and flowers. The **root meristem** is located at the opposite side and creates the root.

A meristem contains an organized group of actively dividing stem cells. As discussed in Chapter 20, stem cells retain the ability to divide and differentiate into multiple cell types. As they grow, meristems produce offshoots of proliferating cells. On a shoot meristem, for example, these offshoots or buds give rise to structures such as leaves and flowers. The organization of a shoot meristem is shown in **Figure 24.22**. It is organized into three areas called the **organizing center,** the **central zone,** and the **peripheral zone.** The role of the organizing center is to ensure the proper organization of the meristem and preserve the correct number of actively dividing stem cells. The central zone is an area where undifferentiated stem cells are always maintained. The peripheral zone contains dividing cells that eventually differentiate into plant structures. For example, the peripheral zone may form a bud, which will produce a leaf or flower.

In *Arabidopsis*, the organization of a shoot meristem is controlled by two critical genes: *WUS* and *CLV3*. The *WUS* gene encodes a transcription factor that is expressed in the organizing center (see Figure 24.22). The expression of the *WUS* gene

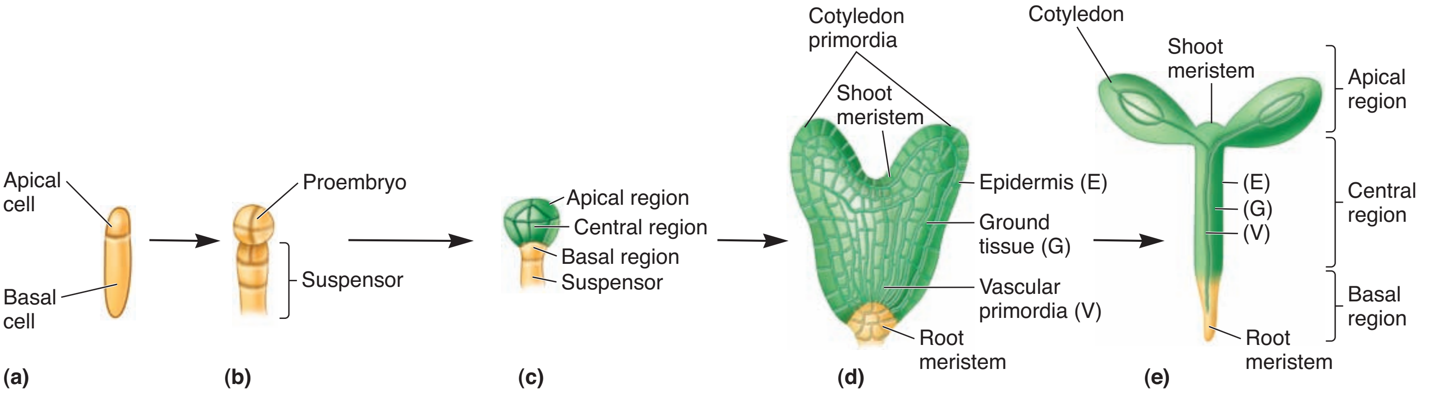


FIGURE 24.21 **Developmental steps in the formation of a plant embryo.** **(a)** The two-cell stage consists of the apical cell and basal cell. **(b)** The eight-cell stage consists of a proembryo and a suspensor. The suspensor gives rise to extraembryonic tissue, which is needed for seed formation. **(c)** At this stage of embryonic development, the three main regions of the embryo (i.e., apical, central, and basal) have been determined. **(d)** At the "heart" stage, all of the plant tissues have begun to form. Note that the shoot meristem is located between the future cotyledons, and the root meristem is on the opposite side. **(e)** A seedling.

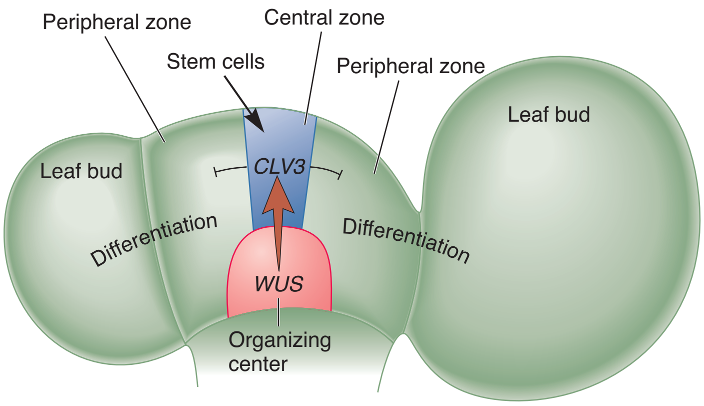


FIGURE 24.22 **Organization of a shoot meristem.** The organization of a shoot meristem is controlled by the *WUS* and *CLV3* genes, which are abbreviations for *Wuschel* and *clavata*, respectively. The *WUS* gene is expressed in the organizing center and induces the cells in the central zone to become undifferentiated stem cells. The red arrow indicates that the WUS protein induces these stem cells to turn on the *CLV3* gene, which encodes a secreted protein that binds to receptors in the cells of the peripheral zone. The black lines with a vertical slash indicate that the CLV3 protein prevents the cells in the peripheral zone from expressing the *WUS* gene. This limits the area of *WUS* gene expression to the underlying organizing center, thereby maintaining a small population of stem cells at the growing tip. The cells in the peripheral zone are allowed to grow and differentiate into lateral structures such as leaves.

Concept check: *Why is it important to maintain the correct number of stem cells in the growing tip?*

induces the adjacent cells in the central zone to become undifferentiated stem cells. These stem cells then turn on the *CLV3* gene, which encodes a secreted protein. The *CLV3* protein binds to receptors in the cells of the peripheral zone, preventing them from expressing the *WUS* gene. This limits the area of *WUS* gene expression to the underlying organizing center, thereby maintaining a small population of stem cells at the growing tip. A shoot meristem in *Arabidopsis* contains only about 100 cells. The inhibition of *WUS* expression in the peripheral cells also allows them to embark on a path of cell differentiation so they can produce structures such as leaves and flowers.

Plant Homeotic Genes Control Flower Development

Although the term *homeotic* was coined by William Bateson to describe homeotic mutations in animals, the first known mutations in homeotic genes were observed in plants. In ancient Greece and Rome, for example, double flowers in which stamens were replaced by petals were noted. In current research, geneticists have been studying these types of mutations to better understand developmental pathways in plants. Many homeotic mutations affecting flower development have been identified in *Arabidopsis* and also in the snapdragon (*Antirrhinum majus*).

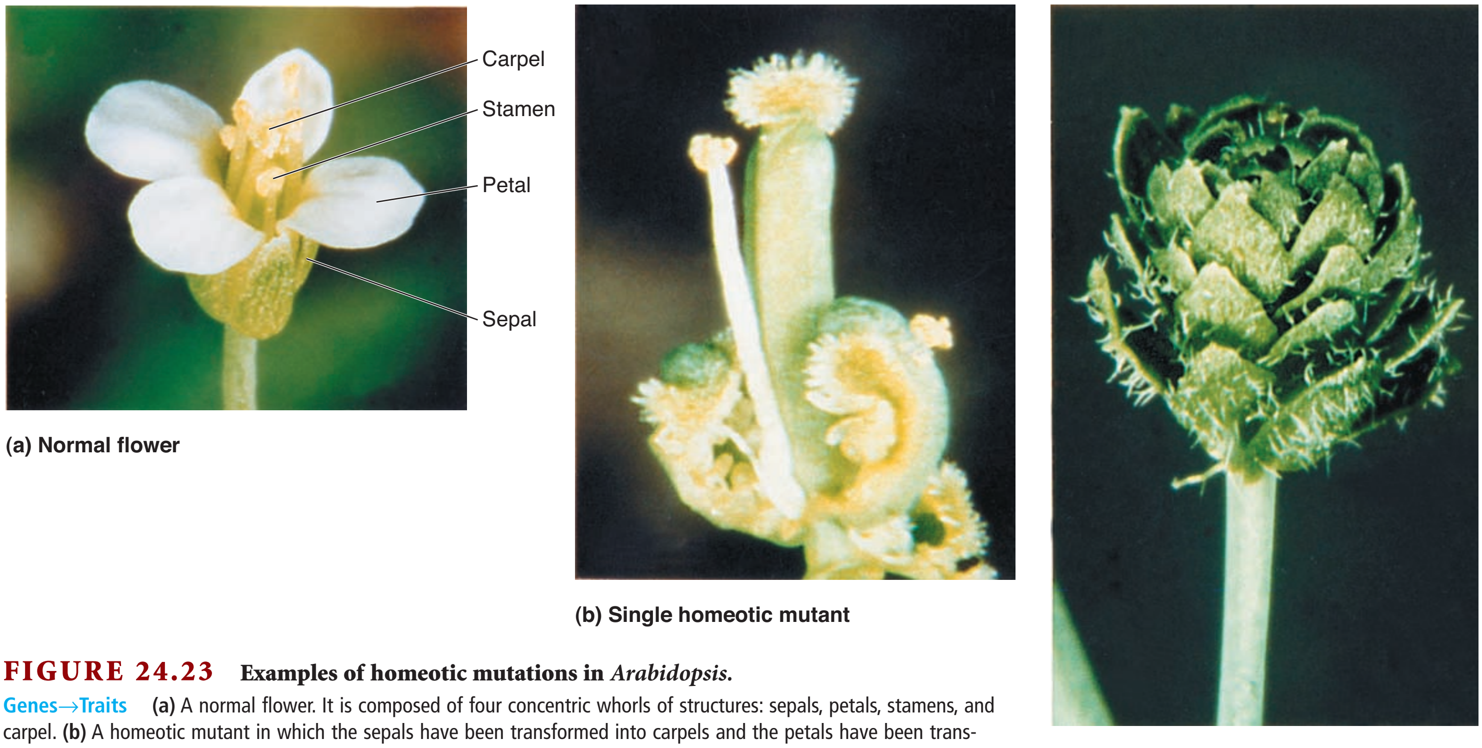
Examples of homeotic mutants in *Arabidopsis* are shown in **Figure 24.23**. A normal *Arabidopsis* flower is composed of four concentric whorls of structures (Figure 24.23a). The outer whorl contains four sepals, which protect the flower bud before it opens. The second whorl is composed of four petals, and the third whorl contains six stamens, the structures that make the male gametophytes, pollen. Finally, the innermost whorl contains two carpels, which are fused together. The carpel produces the female gametophytes. The homeotic mutants shown in Figure 24.23 have undergone transformations of particular whorls. For example, in Figure 24.23b, the sepals have been transformed into carpels, and the petals into stamens.

By analyzing the effects of many different homeotic mutations in *Arabidopsis*, Elliot Meyerowitz and colleagues proposed the **ABC model** for flower development. In this model, three classes of genes, called *A*, *B*, and *C*, govern the formation of sepals, petals, stamens, and carpels. More recently, a fourth category of genes called the *Sepallata* genes (*SEP* genes) have been found to be required for this process. **Figure 24.24** illustrates how these genes affect normal flower development in *Arabidopsis*. Gene *A* products are made in the outermost whorl (whorl 1), promoting sepal formation. In whorl 2, gene *A*, gene *B*, and *SEP* gene products are made, which promotes petal formation. The expression of gene *B*, gene *C*, and *SEP* genes in whorl 3 causes stamens to be made. Finally, in whorl 4, gene *C* and *SEP* genes promote carpel formation.

What is the molecular explanation for how the homeotic mutants in the *A*, *B*, *C*, or *SEP* genes cause their phenotypic effects? In the original ABC model, it was proposed that genes *A* and *C* repress each other's expression, and gene *B* functions independently. In a mutant defective in gene *A* expression, gene *C* is also be expressed in whorls 1 and 2. This produces a carpel-stamen-stamen-carpel arrangement. When gene *B* is defective, a flower cannot make petals or stamens. Therefore, a gene *B* defect yields a flower with a sepal-sepal-carpel-carpel arrangement. When gene *C* is defective, gene *A* is expressed in all four whorls. This results in a sepal-petal-petal-sepal pattern. If the expression of *SEP* genes is defective, the flower consists entirely of sepals, which is the origin of the gene's name.

Overall, the types of genes described in Figure 24.24 promote cell differentiation that leads to sepal, petal, stamen, or carpel structures. But what happens if genes *A*, *B*, and *C* are all defective? As shown in Figure 24.23c, this produces a "flower" that is composed entirely of leaves! These results indicate that the leaf structure is the default pathway and that the *A*, *B*, *C*, and *SEP* genes cause development to deviate from a leaf structure in order to make something else. In this regard, the sepals, petals, stamens, and carpels can be viewed as modified leaves. Interestingly, the German philosopher and poet Johann Goethe originally proposed this idea—that flower formation comes from modifications of the leaf—over 200 years ago.

Arabidopsis has two types of gene *A* (*apetala1* and *apetala2*), two types of gene *B* (*apetala3* and *pistillata*), one type of gene *C* (*agamous*), and three *SEP* genes (*SEP1*, *SEP2*, and *SEP3*). All of these plant homeotic genes encode transcription factor proteins that contain a DNA-binding domain and a dimerization domain. However, the *Arabidopsis* homeotic genes do not contain a sequence similar to the homeobox found in animal homeotic genes.



(a) Normal flower

(b) Single homeotic mutant

(c) Triple mutant

FIGURE 24.23 **Examples of homeotic mutations in *Arabidopsis.***

Genes→Traits **(a)** A normal flower. It is composed of four concentric whorls of structures: sepals, petals, stamens, and carpel. **(b)** A homeotic mutant in which the sepals have been transformed into carpels and the petals have been transformed into stamens. **(c)** A triple mutant in which all of the whorls have been changed into leaves.

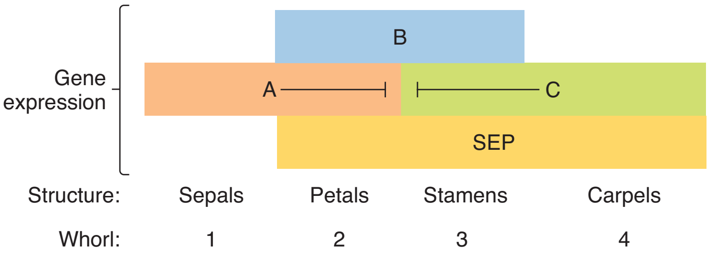


FIGURE 24.24 **The ABC model of homeotic gene action in *Arabidopsis.*** Note: This is a revised model based on the identification of *SEP* genes. The black lines with a vertical slash indicate that the *A* gene product represses the *C* gene product, and vice versa.

Concept check: What would be the expected result if the A gene was inactive?

Like the *Drosophila* homeotic genes, plant homeotic genes are part of a hierarchy of gene regulation. Genes that are expressed within a flower bud produce proteins that activate the expression of these homeotic genes. Once they are transcriptionally activated, the homeotic genes then regulate the expression of other genes, the products of which promote the formation of sepals, petals, stamens, or carpels.

24.4 REVIEWING THE KEY CONCEPTS

- *Arabidopsis thaliana* is a model organism for studying plant development (see Figure 24.20).
- Plant growth occurs from shoot and root meristems (see Figure 24.21).
- The expression of *WUS* and *CLV3* controls the correct number of stem cells in the central zone of a shoot meristem (see Figure 24.22).
- The ABC model describes how homeotic genes control flower development in plants (see Figures 24.23, 24.24).

24.4 COMPREHENSION QUESTIONS

1. The growth of plants is due to the division of __________, which are found in __________.
   a. stem cells, the shoots
   b. stem cells, apical and basal meristems
   c. somatic cells, the shoots
   d. somatic cells, apical and basal meristems
2. Flower development occurs in __________ according to __________.
   a. 3 whorls, maternal effect genes
   b. 3 whorls, the ABC model
   c. 4 whorls, maternal effect genes
   d. 4 whorls, the ABC model

24.5 IMMUNOGENETICS

Antibodies, or **immunoglobulins (Igs),** are proteins produced by the immune system of vertebrates. Their function is to recognize foreign material, such as viruses and bacteria, and target them for destruction. A foreign substance that elicits an immune response is called an **antigen.** Antibodies recognize sites within antigens. The recognition between an antibody and antigen is very specific.

B cells, a type of lymphocyte, are responsible for producing antibodies within the immune system of mammals. Each B cell produces a single type of antibody, and a single individual can produce millions of B cells. Genetic changes to the genes that encode antibody polypeptides allow great variation in the amino acid sequences of the antibodies that are produced by B cells. These differences in the amino acid sequences of antibody proteins enable them to recognize different antigens. In this way, the immune system can identify an impressive variety of foreign materials as being antigens and thereby target them for destruction. In this section, we will consider how this diversity arises.

Antibody Diversity in the Immune System Is Produced by Site-Specific Recombination

From a genetic viewpoint, the production of millions of different antibodies poses an interesting problem. If a distinct gene was needed to produce each different antibody polypeptide, the genome would need to contain millions of different antibody genes. By comparison, consider that the entire human genome contains only about 20,000 to 25,000 different genes. How then is it possible for the genome to produce millions of different antibody proteins, which are encoded by genes? The answer to this question baffled geneticists for several decades until research revealed that antibodies with different polypeptide sequences are generated by an unusual mechanism in which the antibody genes are cut and reconnected by **site-specific recombination.** Within a large population of B cells, these genes are cut and rejoined in many different ways to produce a vast array of polypeptides with differing amino acid sequences. With this mechanism, only a few antibody genes are needed to produce millions of different antibodies.

Antibodies are tetrameric proteins composed of two heavy polypeptide chains and two light chains (**Figure 24.25**). The sequences of amino acids within the variable regions vary among different antibodies. Because the variable region is the site that recognizes an antigen, this variation allows the immune system to recognize many different types of antigens.

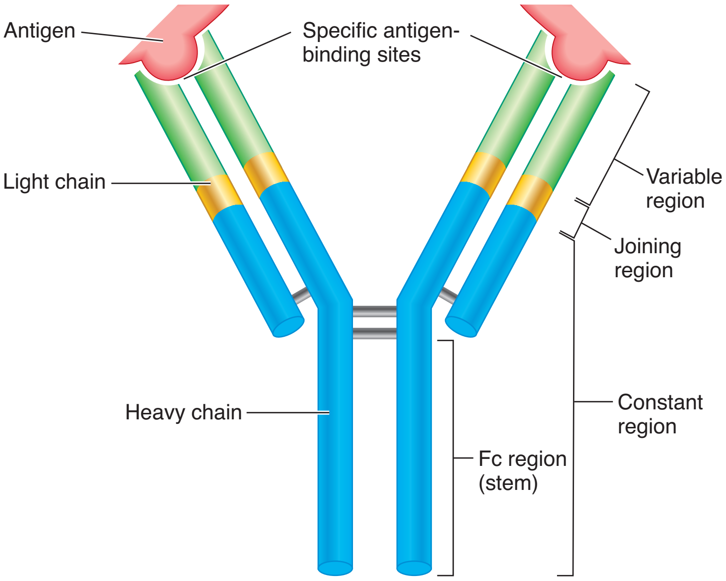


FIGURE 24.25 **Structure of an antibody.**

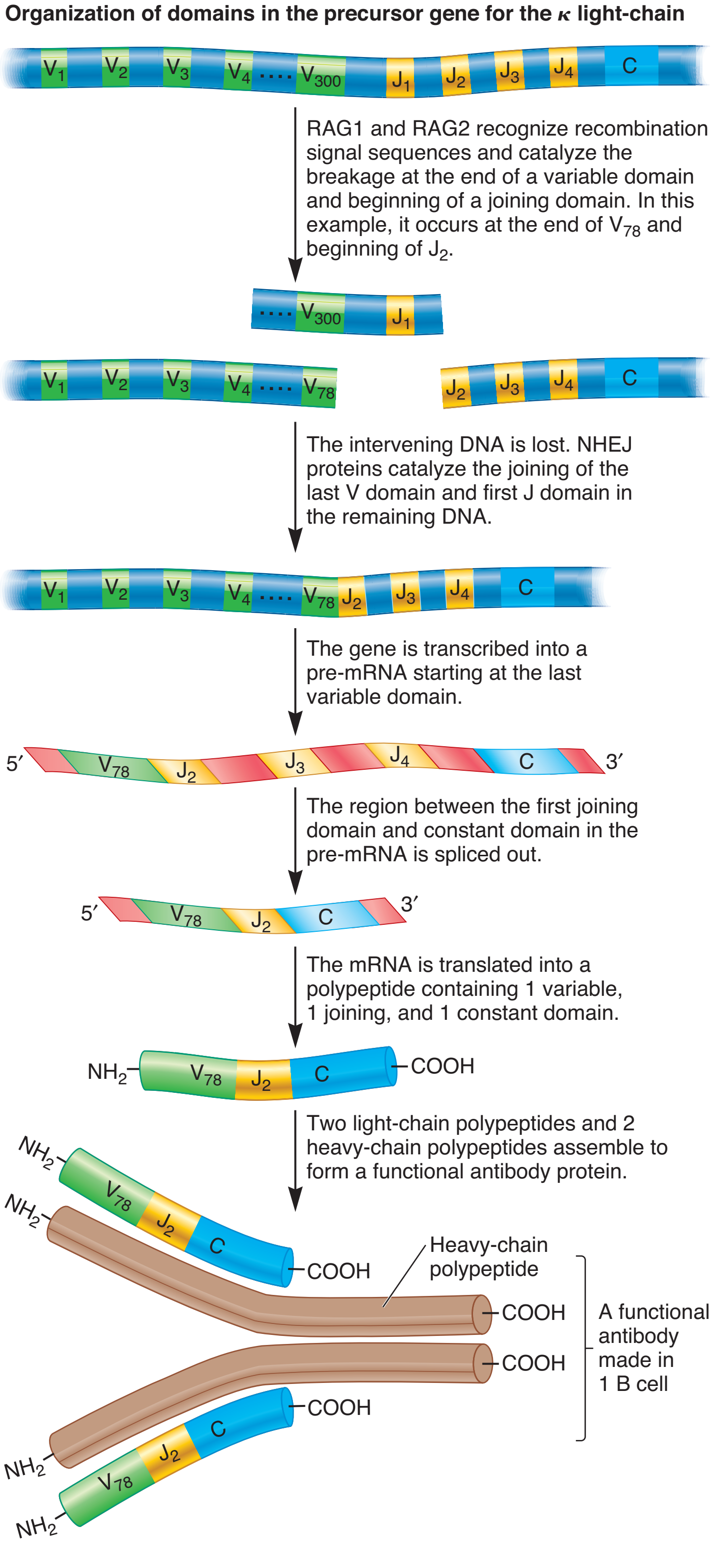
One type of light chain is the κ (kappa) light chain, which is found in different classes of antibodies such as immunoglobulin G (IgG). Mammals have multiple copies of the gene that encodes the light and heavy chains of antibodies. As an example, the top of **Figure 24.26** shows the organization of one precursor gene for the κ light chain that is found in the mouse. This type of gene is sometimes called a precursor gene because its structure is subsequently altered by site-specific recombination. As shown on the upper left side of Figure 24.26, the precursor gene has approximately 300 regions known as variable (V) sequences or domains. In addition, the gene contains four different joining (J) sequences and a single constant (C) sequence. Each variable domain or joining domain encodes a different amino acid sequence.

Figure 24.26 shows the site-specific recombination of a light-chain precursor gene in the mouse. During the maturation of B cells, the κ light-chain precursor gene is cut and rejoined so that one variable domain becomes adjacent to a joining domain. At the end of every V domain and the beginning of every J domain is located a **recombination signal sequence** that functions as a recognition site for site-specific recombination between the V and J regions. The recombination event is initiated by two proteins called **RAG1** and **RAG2.** RAG is an acronym for recombination-activating gene. These proteins recognize recombination signal sequences and generate two double-strand breaks: one at the end of a V domain and one at the beginning of a J domain. For example, in the recombination event shown in Figure 24.26, RAG1 and RAG2 have made cuts at the end of variable domain number 78 and the beginning of joining domain number 2. The intervening region of DNA is lost, and the two ends are then joined to each other. The connection phase of this process is catalyzed by a group of proteins termed **nonhomologous end-joining (NHEJ) proteins.**

Following transcription, the fused VJ region is contained within a pre-mRNA transcript that is then spliced to connect the J and C domains. After this has occurred, each B cell produces only the particular κ light chain encoded by the specific VJ fusion domain and the constant domain.

The heavy-chain polypeptides are produced by a similar recombination mechanism. Though mammals have multiple copies of heavy-chain genes, a typical gene in mice may have about 500 variable domains and four joining domains. In addition, a heavy-chain gene encodes several diversity (D) domains, which are found between the variable and joining domains. The recombination first involves the connection of a D and J domain, followed by the connection of a V and DJ domain. The same proteins that catalyze VJ fusion are involved in the recombination with the heavy-chain gene. Collectively, this process is called **V(D)J recombination.** The D is in parentheses because this type of domain is found only in the heavy-chain genes, not in the light-chain genes.

The recombination process within immunoglobulin genes produces an enormous diversity in polypeptides. Even though it


Organization of domains in the precursor gene for the κ light-chain

RAG1 and RAG2 recognize recombination signal sequences and catalyze the breakage at the end of a variable domain and beginning of a joining domain. In this example, it occurs at the end of V_{78} and beginning of J_2.

The intervening DNA is lost. NHEJ proteins catalyze the joining of the last V domain and first J domain in the remaining DNA.

The gene is transcribed into a pre-mRNA starting at the last variable domain.

The region between the first joining domain and constant domain in the pre-mRNA is spliced out.

The mRNA is translated into a polypeptide containing 1 variable, 1 joining, and 1 constant domain.

Two light-chain polypeptides and 2 heavy-chain polypeptides assemble to form a functional antibody protein.

A functional antibody made in 1 B cell

FIGURE 24.26 **Site-specific recombination within the antibody precursor gene that encodes the κ light chain for immunoglobulin G (IgG) proteins.**

Concept check: What is the purpose of site-specific recombination within the antibody precursor gene?

occurs at specific junctions within the antibody gene, the recombination is fairly random with regard to the particular V and J domains that are joined in any given B cell.

Overall, the possible number of functional antibodies that can be produced by V(D)J recombination is rather staggering. For example, if we assume that any of the 300 different variable sequences can be spliced next to any of the four joining sequences, this results in 1200 possible light-chain combinations. If we also assume that a heavy-chain gene has 500 variable regions, 12 diversity domains, and 4 joining regions, the number of heavy-chain possibilities is 500 × 12 × 4 = 24,000. Because any light-chain/heavy-chain combination is possible, this yields 1200 × 24,000 = 28,800,000 possible antibody molecules from the random recombination within two precursor genes! The diversity is actually higher because mammals have multiple copies of nonidentical precursor genes for the light and heavy chains of antibodies.

Antibody Diversity Is Also Produced by Imprecise End Joining and Somatic Hypermutation

As we have seen, the rearrangement of V(D)J domains within antibody genes is one mechanism of generating antibody diversity. Two other mechanisms also play a role. As mentioned, the connection phase of this gene rearrangement process is catalyzed by NHEJ proteins. The process may not be entirely precise, so a few nucleotides can be added or lost at the junction between the variable and joining domains. This imprecision is a second mechanism that further accentuates the diversity in antibody genes.

A third mechanism that generates antibody diversity is **somatic hypermutation,** which means that antibody genes undergo a high rate of mutation in somatic cells (in this case, B cells). During this process, cytosines within antibody genes are deaminated into uracils. After DNA replication, this event leads to numerous C to T mutations. In addition, the presence of uracils recruits lesion-replicating DNA polymerases (see Chapter 13). These DNA polymerases are error-prone, which leads to several additional types of mutations. The end result of somatic hypermutation is the addition of new mutations within the antibody genes, thereby generating additional diversity.

24.5 REVIEWING THE KEY CONCEPTS

- Antibodies are tetrameric proteins that recognize antigens (see Figure 24.25).
- Antibody precursor genes are rearranged via site-specific recombination to produce a vast array of different antibodies that can recognize different antigens (see Figure 24.26).
- Imprecise end joining and somatic hypermutation also change the sequences of antibody genes.

24.5 COMPREHENSION QUESTIONS

1. Antibody diversity is produced by
   a. V(D)J recombination.
   b. imprecise end joining during V(D)J recombination.
   c. somatic hypermutation.
   d. all of the above.

KEY TERMS

Page 585. developmental genetics

Page 586. pattern, positional information, apoptosis, morphogen

Page 587. induction, cell adhesion, cell adhesion molecules (CAMs), homeotic gene

Page 588. bilaterians, determination

Page 589. segments

Page 591. embryogenesis, anteroposterior axis, dorsoventral axis, left-right axis, proximodistal axis

Page 592. maternal effect

Page 593. segmentation genes

Page 594. zygotic genes

Page 595. cell fate

Page 596. homeotic, *Hox* genes, gain-of-function mutation

Page 597. homeobox, homeodomain, cell lineage diagram

Page 598. cell lineage

Page 599. homologous genes, orthologs, *Hox* complexes, reverse genetics, gene knockout

Page 600. determined, differentiated

Page 601. basic domain, helix-loop-helix domain (bHLH), myogenic bHLH proteins

Page 602. totipotent

Page 603. shoot meristem, root meristem, organizing center, central zone, peripheral zone

Page 604. ABC model

Page 605. antibodies, immunoglobulins (Igs), antigen

Page 606. site-specific recombination, recombination signal sequence, RAG1, RAG2, nonhomologous end-joining (NHEJ) proteins, V(D)J recombination

Page 607. somatic hypermutation

CHAPTER SUMMARY

- Developmental genetics is concerned with the roles that genes play in orchestrating the changes that occur during development.

24.1 Overview of Animal Development

- Positional information in animals may cause a cell to divide, migrate, differentiate, or undergo apoptosis (see Figure 24.1).
- A morphogen is a molecule that conveys positional information. Three molecular mechanisms that convey positional information include asymmetrical distribution of maternally deposited morphogens in an oocyte, asymmetrical synthesis and extracellular distribution of a morphogen in a developing embryo, and cell-to-cell contact (see Figure 24.2).
- The study of mutants that disrupt development has identified genes that control development (see Figure 24.3).
- Development in animals involves four overlapping phases: the formation of axes, segmentation, determination, and cell differentiation (see Figure 24.4).

24.2 Invertebrate Development

- *Drosophila* proceeds through different developmental stages from a fertilized egg to an adult. Various sets of genes are responsible for developmental changes (see Figure 24.5, Table 24.1).
- Maternal effect genes establish the anteroposterior and dorsoventral axes due to their asymmetrical distribution. An example is *bicoid*, which promotes the formation of anterior structures (see Figures 24.6–24.8).
- The *Drosophila* embryo is divided into segments (see Figure 24.9).
- Three categories of segmentation genes have been identified based on analyzing the phenotypes of mutations affecting segmentation (see Figure 24.10).
- A hierarchy of gene expression, which includes maternal effect genes, gap genes, pair-rule genes, and segment-polarity genes, gives rise to a segmented *Drosophila* embryo (see Figure 24.11).
- Homeotic genes control the developmental fate of particular segments (see Figures 24.12, 24.13).
- Homeotic proteins contain a DNA-binding domain and a transcriptional activation domain (see Figure 24.14).
- In *C. elegans*, a cell lineage diagram describes the cell fate of each cell of the worm's body (see Figure 24.15).

24.3 Vertebrate Development

- Homeotic genes in vertebrates are found in *Hox* complexes (see Figure 24.16).
- *Hox* genes control the fate of regions along the anteroposterior axis (see Figures 24.17, 24.18).
- Transcription factors also control cell differentiation. An example is *MyoD*, which causes cells to differentiate into skeletal muscle cells (see Figure 24.19).

24.4 Plant Development

- *Arabidopsis thaliana* is a model organism for studying plant development (see Figure 24.20).
- Plant growth occurs from shoot and root meristems (see Figure 24.21).
- The expression of *WUS* and *CLV3* controls the correct number of stem cells in the central zone of a shoot meristem (see Figure 24.22).
- The ABC model describes how homeotic genes control flower development in plants (see Figures 24.23, 24.24).

24.5 Immunogenetics

- Antibodies are tetrameric proteins that recognize antigens (see Figure 24.25).
- Antibody precursor genes are rearranged via site-specific recombination to produce a vast array of different antibodies that can recognize different antigens (see Figure 24.26).
- Imprecise end joining and somatic hypermutation also change the sequences of antibody genes.

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Discuss the functional roles of the maternal effect genes, gap genes, pair-rule genes, and segment-polarity genes in *Drosophila*.

Answer: These genes are involved in pattern formation of the *Drosophila* embryo. The asymmetrical distribution of maternal effect gene products in the oocyte establishes the anteroposterior and dorsoventral axes. These gene products also control the expression of the gap genes, which are expressed as broad bands in certain regions of the embryo. The overlapping expression of maternal effect genes and gap genes controls the pair-rule genes, which are expressed in alternating stripes. A stripe corresponds to a parasegment. Within each parasegment, the expression of segment-polarity genes defines an anterior and posterior compartment. With regard to morphology, an anterior compartment of one parasegment and the posterior compartment of an adjacent parasegment form a segment of the fly.

S2. With regard to genes affecting development, what are the phenotypic effects of gain-of-function mutations versus loss-of-function mutations?

Answer: Gain-of-function mutations cause a gene to be expressed in the wrong place, at the wrong time, or at an abnormally high level. When the gene is expressed in the wrong place, that region may develop into an inappropriate structure. For example, when *Antp* is abnormally expressed in an anterior segment, this segment develops legs in place of antennae. By comparison, loss-of-function mutations result in a defect in the expression of a gene with the result that protein function is reduced or abolished. This usually disrupts the developmental process, because the cells in the region where the gene is normally expressed are not directed to develop along the correct pathway.

S3. Mutations in genes that control the early stages of development are often lethal (e.g., see Figure 24.7b). To circumvent this problem, developmental geneticists may try to isolate temperature-sensitive developmental mutants, or *ts* alleles. If an embryo carries a *ts* allele, it will develop correctly at the permissive temperature (e.g., 25°C) but will fail to develop if incubated at the nonpermissive temperature (e.g., 30°C). *Ts* alleles are particularly useful because they can provide insight regarding the stage of development when the protein encoded by the gene is necessary. Researchers can take groups of embryos that carry a *ts* allele and expose them to the permissive and nonpermissive temperatures at different stages of development. In the experiment described next, embryos were divided into five groups and exposed to the permissive or nonpermissive temperature at different times after fertilization.

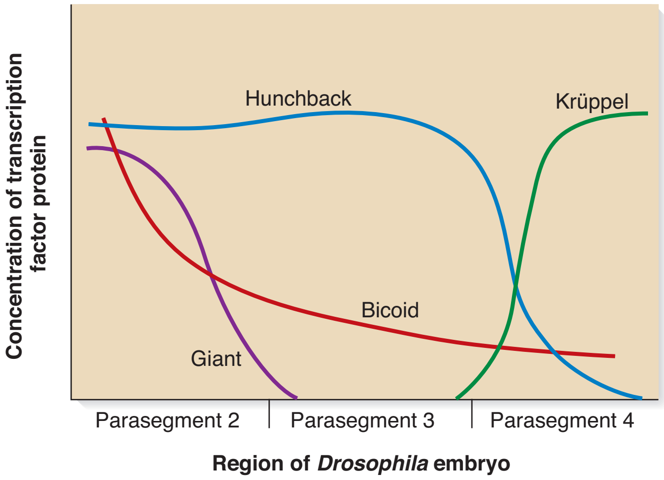
Time After Fertilization (hours)	**Group: 1**	**2**	**3**	**4**	**5**
0–1	25°C	25°C	25°C	25°C	25°C
1–2	25°C	30°C	25°C	25°C	25°C
2–3	25°C	25°C	30°C	25°C	25°C
3–4	25°C	25°C	25°C	30°C	25°C
4–5	25°C	25°C	25°C	25°C	30°C
5–6	25°C	25°C	25°C	25°C	25°C
SURVIVAL:	Yes	Yes	Yes	No	Yes

Explain these results.

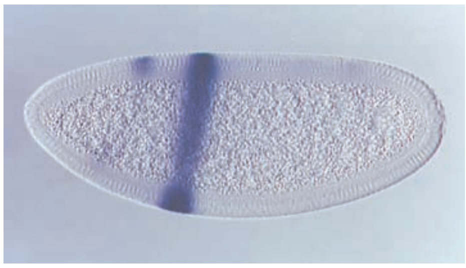
Answer: By varying the temperature during different stages of development, researchers can pinpoint the stage when the function of the protein encoded by this *ts* allele is critical. As shown, embryos fail to survive if they are subjected to the nonpermissive temperature between 3 and 4 hours after fertilization, but they do survive if subjected to the nonpermissive temperature at other times of development. These results indicate that this protein plays a crucial role at the 3- to 4-hour stage of development.

S4. An intriguing question in developmental genetics is, How can a particular gene, such as *even-skipped*, be expressed in a multiple banding pattern as seen in Figure 24.11? Another way of asking this question is, How is the positional information within the broad bands of the gap genes deciphered in a way that causes the pair-rule genes to be expressed in this alternating banding pattern? The answer lies in a complex mechanism of genetic regulation. Certain pair-rule genes have several stripe-specific enhancers that are controlled by multiple transcription factors. A stripe-specific enhancer is typically a short segment of DNA, 300 to 500 bp in length, that contains binding sequences recognized by several different transcription factors. This term is a bit misleading because a stripe-specific enhancer is a regulatory region that contains both enhancer and silencer elements.

In 1992, Michael Levine and his colleagues investigated stripe-specific enhancers that are located near the promoter of the *even-skipped* gene. A segment of DNA, termed the stripe 2 enhancer, controls the expression of the *even-skipped* gene; this enhancer is responsible for the expression of the *even-skipped* gene in stripe 2, which corresponds to parasegment 3 of the embryo. The stripe 2 enhancer is a segment of DNA that contains binding sites for four transcription factors that are the products of the *Krüppel*, *bicoid*, *hunchback*, and *giant* genes. The Hunchback and Bicoid transcription factors bind to this enhancer and activate the transcription of the *even-skipped* gene. In contrast, the transcription factors encoded by the *Krüppel* and *giant* genes bind to the stripe 2 enhancer and repress transcription. The figure shown here describes the concentrations of these four transcription factor proteins in the region of parasegment 3 (i.e., stripe 2) in the *Drosophila* embryo.



To study stripe-specific enhancers, researchers have constructed artificial genes in which the enhancer is linked to a reporter gene, the expression of which is easy to detect. The next figure shows the results of an experiment in which an artificial gene was made by placing the stripe 2 enhancer next to the β-galactosidase gene. This artificial gene was introduced into *Drosophila*, and then embryos containing this gene were analyzed for β-galactosidase activity. If a region of the embryo is expressing β-galactosidase, the region will stain darkly because β-galactosidase converts a colorless compound into a dark blue compound.



Explain these results.

Answer: As shown in the first figure to this problem, the concentrations of the Hunchback and Bicoid transcription factors are relatively high in the region of the embryo corresponding to stripe 2 (which is parasegment 3). The levels of Krüppel and Giant are very low in this region. Therefore, the high levels of activators and low levels of repressors cause the *even-skipped* gene to be transcribed. In this experiment, β-galactosidase was made only in stripe 2 (i.e., parasegment 3). These results show that the stripe-2-specific enhancer controls gene expression only in parasegment 3. Because we know that the *even-skipped* gene is expressed as several alternating stripes (as seen in Figure 24.11), the *even-skipped* gene must contain other stripe-specific enhancers that allow it to be expressed in these other alternating parasegments.

Conceptual Questions

C1. In the case of multicellular animals, what are the four types of cellular processes that occur so a fertilized egg can develop into an adult organism? Briefly discuss the role of each process.

C2. The arrangement of body axes of the fruit fly are shown in Figure 24.5g. Are the following statements true or false with regard to body axes in the mouse?

A. Along the anteroposterior axis, the head is posterior to the tail.

B. Along the dorsoventral axis, the vertebrae of the back are dorsal to the stomach.

C. Along the dorsoventral axis, the feet are dorsal to the hips.

D. Along the proximodistal axis, the toes on the hind legs are distal to the hips of the hind legs.

C3. If you observed fruit flies with the following developmental abnormalities, would you guess that a mutation has occurred in a segmentation gene or a homeotic gene? Explain your guess.

A. Three abdominal segments were missing.

B. One abdominal segment had legs.

C. A fly with the correct number of segments had two additional thoracic segments and two fewer abdominal segments.

C4. Which of the following statements are true with regard to positional information in *Drosophila*?

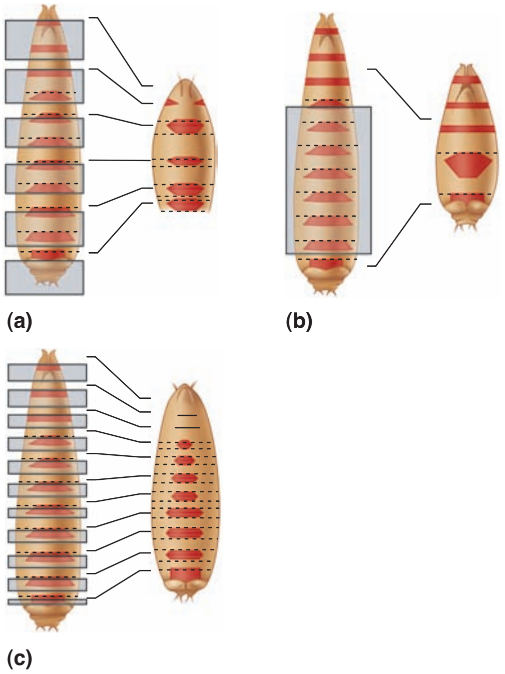
A. Morphogens are a type of molecule that conveys positional information.

B. Morphogenetic gradients are established only in the oocyte, prior to fertilization.

C. Cell adhesion molecules also provide a way for a cell to obtain positional information.

C5. Discuss the morphological differences between the parasegments and segments of *Drosophila*. Discuss the evidence, providing specific examples, that suggests the parasegments of the embryo are the subdivisions for the organization of gene expression.

C6. Here are schematic diagrams of mutant larvae.



The left side of each pair shows a wild-type larva, with gray boxes showing the sections that are missing in the mutant larva. Which type of gene is defective in each larva: a gap gene, a pair-rule gene, or a segment-polarity gene?

C7. Describe what a morphogen is and how it exerts its effects. What do you expect will happen when a morphogen is expressed in the wrong place in an embryo? List five examples of morphogens that function in *Drosophila*.

C8. What is the meaning of positional information? Discuss three different ways that cells can obtain positional information. Which of these three ways do you think is the most important for the formation of a segmented body pattern in *Drosophila*?

C9. Gradients of morphogens can be preestablished in the oocyte. Also, later in development, morphogens can be secreted from cells. How are these two processes similar and different?

C10. Discuss how the anterior portion of the anteroposterior axis is established. What aspects of oogenesis are critical in establishing this axis? What do you think would happen if the *bicoid* mRNA was not trapped at the anterior end but instead diffused freely throughout the oocyte?

C11. Describe the function of the Bicoid protein. Explain how its ability to exert its effects in a concentration-dependent manner is a critical feature of its function.

C12. With regard to development, what are the roles of the maternal effect genes versus the zygotic genes? Which types of genes are needed earlier in the development process?

C13. Discuss the role of homeotic genes in development. Explain what happens to the phenotype of a fly when a gain-of-function homeotic gene mutation causes the protein to be expressed in an abnormal region of the embryo. What are the consequences of a loss-of-function mutation in such a gene?

C14. Describe the molecular features of the homeobox and homeodomain. Explain how these features are important in the function of homeotic genes.

C15. What would you predict to be the phenotype of a *Drosophila* larva whose mother was homozygous for a loss-of-function allele in the *nanos* gene?

C16. Based on the photographs in Figure 24.13, in which segments would the *Antp* gene normally be expressed?

C17. If a mutation in a homeotic gene produced the following phenotypes, would you expect the mutation to be a loss-of-function or a gain-of-function allele? Explain your answer.

A. An abdominal segment has antennae attached to it.

B. The most anterior abdominal segment resembles the most posterior thoracic segment.

C. The most anterior thoracic segment resembles the most posterior abdominal segment.

C18. Explain how loss-of-function mutations in the following categories of genes would affect the morphologies of *Drosophila* larvae:

A. Gap genes

B. Pair-rule genes

C. Segment-polarity genes

C19. What is the difference between a maternal effect gene and a zygotic gene? Of the following genes that play a role in *Drosophila* development, which would be maternal effect genes and which would be zygotic? Explain your answer.

A. *nanos* C. *bicoid*

B. *Antp* D. *lab*

C20. Cloning of mammals (such as Dolly) is described in Chapter 20. Based on your understanding of animal development, explain why an enucleated egg is needed to clone mammals. In other words, what features of the oocyte are essential for animal development?

C21. Discuss the similarities and differences between the *bithorax* and *Antennapedia* complexes in *Drosophila* and the *Hox* gene complexes in mice.

C22. What is cell differentiation? Discuss the role of bHLH proteins in the differentiation of muscle cells. Explain how they work at the molecular level. In your answer, indicate how protein dimerization is a key feature in gene regulation.

C23. The *MyoD* gene in mammals plays a role in muscle cell differentiation, whereas the *Hox* genes are homeotic genes that play a role in the differentiation of particular regions of the body. Explain how the functions of these genes are similar and different.

C24. What is a totipotent cell? In the following examples, which cells would be totipotent? In the case of multicellular organisms, such as humans and corn, explain which cell types would be totipotent.

A. Humans C. Yeast

B. Corn D. Bacteria

C25. What is a meristem? Explain the role of meristems in plant development.

C26. Discuss the morphological differences between animal and plant development. How are they different at the cellular level? How are they similar at the genetic level?

C27. Predict the phenotypic consequences of each of the following mutations:

A. *apetala1* defective

B. *pistillata* defective

C. *apetala1* and *pistillata* defective

C28. Briefly describe three ways that antibody diversity is produced.

C29. Describe the function of RAG1 and RAG2 proteins and NHEJ proteins.

C30. According to the scenario shown in Figure 24.26, how many segments of DNA (one, two, or three) are removed during site-specific recombination within the gene that encodes the κ (kappa) light chain for IgG proteins? How many segments are spliced out of the pre-mRNA?

Application and Experimental Questions

E1. Researchers have used the cloning methods described in Chapter 19 to clone the *bicoid* gene and express large amounts of the Bicoid protein. The Bicoid protein was then injected into the posterior end of a zygote immediately after fertilization. What phenotypic results would you expect to occur? What do you think would happen if the Bicoid protein was injected into a segment of a larva?

E2. Compare and contrast the experimental advantages of *Drosophila* and *C. elegans* in the study of developmental genetics.

E3. What is meant by the term *cell fate*? What is a cell lineage diagram? Discuss the experimental advantage of having a cell lineage diagram. What is a cell lineage?

E4. Take a look at Solved Problem S3 before answering this question. *Drosophila* embryos carrying a *ts* mutation were exposed to the permissive (25°C) or nonpermissive (30°C) temperature at different stages of development (see next page). Explain these results.

Time After Fertilization (hours)	**Group: 1**	**2**	**3**	**4**	**5**
0–1	25°C	25°C	25°C	25°C	25°C
1–2	25°C	30°C	25°C	25°C	25°C
2–3	25°C	25°C	30°C	25°C	25°C
3–4	25°C	25°C	25°C	30°C	25°C
4–5	25°C	25°C	25°C	25°C	30°C
5–6	25°C	25°C	25°C	25°C	25°C
SURVIVAL:	Yes	No	No	Yes	Yes

E5. All of the homeotic genes in *Drosophila* have been cloned. As discussed in Chapter 19, cloned genes can be manipulated in vitro. They can be subjected to cutting and pasting, site-directed mutagenesis, etc. After *Drosophila* genes have been altered in vitro, they can be inserted into a *Drosophila* transposon vector (i.e., a P element vector), and then the genetic construct containing the altered gene within a P element can be injected into *Drosophila* embryos. The P element then transposes into the chromosomes, thereby introducing one or more copies of the altered gene into the *Drosophila* genome. This method is termed *P element transformation.*

With these ideas in mind, how would you make a mutant gene with a "gain-of-function" in which the *Antp* gene would be expressed where the *abd-A* gene is normally expressed? What phenotype would you expect for flies that carried this altered gene?

E6. You need to understand Solved Problem S4 before answering this question. If the artificial gene containing the stripe 2 enhancer and the β-galactosidase gene was found within an embryo that also contained the following loss-of-function mutations, what results would you expect? In other words, would there be a stripe or not? Explain why.

A. *Krüppel*

B. *bicoid*

C. *hunchback*

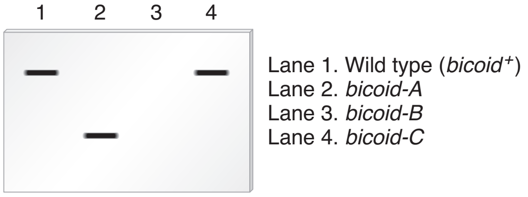
D. *giant*

E7. Two techniques commonly used to study the expression patterns of genes that play a role in development are Northern blotting and fluorescence in situ hybridization (FISH). As described in Chapter 19, Northern blotting can be used to detect RNA that is transcribed from a particular gene. In this method, a specific RNA is detected by using a short segment of cloned DNA as a probe. The DNA probe, which is radioactive, is complementary to the RNA that the researcher wishes to detect. After the radioactive probe DNA binds to the RNA within a blot of a gel, the RNA is visualized as a dark (radioactive) band on an X-ray film. For example, a DNA probe that is complementary to the *bicoid* mRNA could be used to specifically detect the amount and size of the mRNA in a blot.

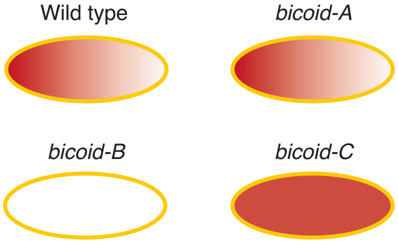
A second technique, termed *fluorescence in situ hybridization (FISH),* can be used to identify the locations of genes on chromosomes. This technique can also be used to locate gene products within oocytes, embryos, and larvae. For this reason, it has been commonly used by developmental geneticists to understand the expression patterns of genes during development. The micrograph in Figure 24.8b is derived from the application of the FISH technique. In this case, the probe was complementary to *bicoid* mRNA.

Now here is the question. Suppose a researcher has three different *Drosophila* strains that have loss-of-function mutations in the *bicoid* gene. We will call them *bicoid-A*, *bicoid-B,* and *bicoid-C;* the wild type is designated *bicoid*$^+$. To study these mutations, phenotypically normal female flies that are homozygous for the *bicoid* mutation were obtained, and their oocytes were analyzed using these two techniques. A wild-type strain was also analyzed as a control. In other words, RNA was isolated from some of the oocytes and analyzed by Northern blotting, and some oocytes were subjected to in situ hybridization. In both cases, the probe was complementary to the *bicoid* mRNA. The results are shown here.

Northern blot



In situ hybridization



A. How can phenotypically normal female flies be homozygous for a loss-of-function allele in the *bicoid* gene?

B. Explain the type of mutation (e.g., deletion, point mutation, etc.) in each of the three strains. Explain how the mutation may cause a loss of normal function for the *bicoid* gene product.

C. Discuss how the use of both techniques provides more definitive information than the application of just one of the two techniques.

E8. Explain one experimental strategy you could follow to determine the functional role of the mouse *HoxD-3* gene.

E9. Another way to study the role of proteins (e.g., transcription factors) that function in development is to microinject the mRNA that encodes a protein, or the purified protein itself, into an oocyte or embryo, and then determine how this affects the subsequent development of the embryo, larva, and adult. For example, if Bicoid protein is injected into the posterior region of an oocyte, the resulting embryo will develop into a larva that has anterior structures at both ends. Based on your understanding of the function of these developmental genes, what would be the predicted phenotype if the following proteins or mRNAs were injected into normal oocytes or embryos?

A. *Nanos* mRNA injected into the anterior end of an oocyte

B. Antp protein injected into the posterior end of an embryo

C. *Toll* mRNA injected the dorsal side of an early embryo

E10. Why have geneticists been forced to use reverse genetics to study the genes involved in vertebrate development? Explain how this strategy differs from traditional genetic analyses like those done by Mendel.

Questions for Student Discussion/Collaboration

1. Compare and contrast the experimental advantages and disadvantages of *Drosophila*, *C. elegans*, mammals, and *Arabidopsis*.
2. It seems that developmental genetics boils down to a complex network of gene regulation. Try to draw how this network is structured for *Drosophila*. How many genes do you think are necessary to describe a complete developmental network for the fruit fly? How many genes do you think are needed for a network to specify one segment? Do you think it is more difficult to identify genes that are involved in the beginning, middle, or end of this network? Knowing what you know about *Drosophila* development, suppose you were trying to identify all of the genes needed for development in a chicken. Would you first try to identify genes necessary for early development, or would you begin by identifying genes involved in cell differentiation?
3. At the molecular level, how do you think a gain-of-function mutation in a developmental gene might cause it to be expressed in the wrong place or at the wrong time? Explain what type of DNA sequence would be altered.

Answers to Comprehension Questions

24.1: e, d, c
24.2: a, d, c, b
24.3: b, b, c
24.4: b, d
24.5: d

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

25

CHAPTER OUTLINE

25.1 Genes in Populations and the Hardy-Weinberg Equation

25.2 Overview of Microevolution

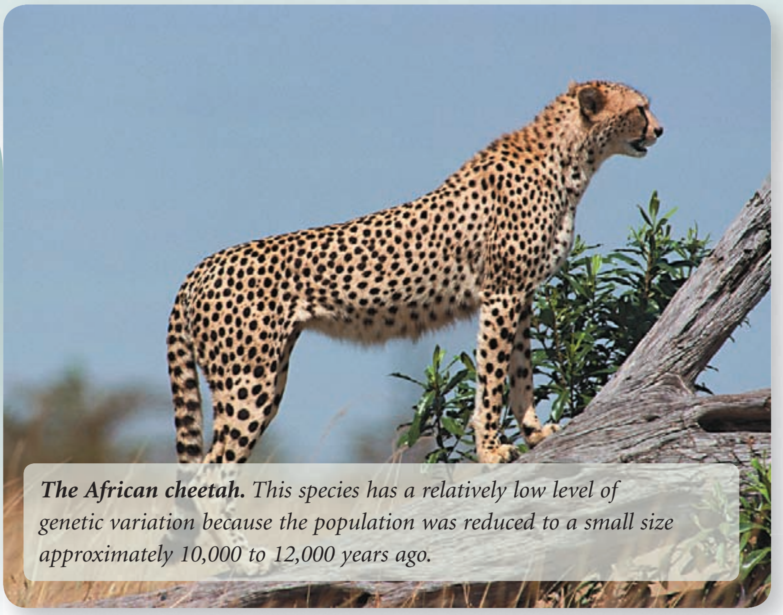
25.3 Natural Selection

25.4 Genetic Drift

25.5 Migration

25.6 Nonrandom Mating

25.7 Sources of New Genetic Variation



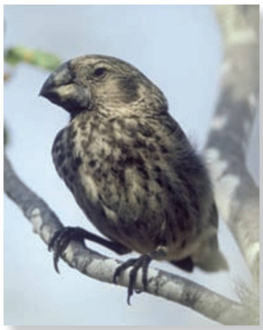
The African cheetah. *This species has a relatively low level of genetic variation because the population was reduced to a small size approximately 10,000 to 12,000 years ago.*

POPULATION GENETICS

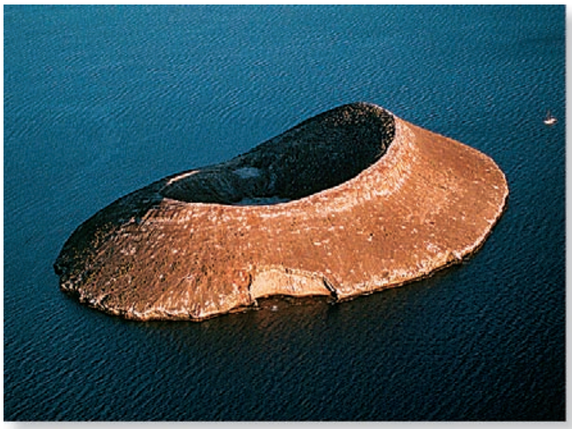
Until now, we have primarily focused our attention on genes within individuals and their related family members. In this chapter and Chapters 26 and 27, we turn to the study of genes in a population or species. The field of **population genetics** is concerned with changes in genetic variation within a group of individuals over time. Population geneticists want to know the extent of genetic variation within populations, why it exists, and how it changes over the course of many generations. The field of population genetics emerged as a branch of genetics in the 1920s and 1930s. Its mathematical foundations were developed by theoreticians who extended the principles of Gregor Mendel and Charles Darwin by deriving formulas to explain the occurrence of genotypes within populations. These foundations can be largely attributed to three scientists: Sir Ronald Fisher, Sewall Wright, and J. B. S. Haldane. As we will see, support for their mathematical theories was provided by several researchers who analyzed the genetic composition of natural and experimental populations. More recently, population geneticists have used techniques to probe genetic variation at the molecular level. In addition, staggering advances in computer technology have aided population geneticists in the analysis of their genetic theories and data. In this chapter, we will explore the genetic variation that occurs in populations and consider the reasons why the genetic composition of populations may change over the course of several generations.

25.1 GENES IN POPULATIONS AND THE HARDY-WEINBERG EQUATION

In the field of population genetics, the focus shifts away from the individual and toward the population of which the individual is a member. Population genetics may seem like a significant departure from other topics in this textbook, but it is a direct extension of our understanding of Mendel's laws of inheritance, molecular genetics, and the ideas of Darwin. Conceptually, all of the alleles of every gene in a population make up the **gene pool.** In this regard, each member of the population is viewed as receiving its genes from its parents, which, in turn, are members of the gene pool. Furthermore, individuals that reproduce contribute to the gene pool of the next generation. Population geneticists study the genetic variation within the gene pool and how such variation changes from one generation to the next. The emphasis is often on allelic variation. In this introductory section, we will examine some of the general features of populations and gene pools.



(a) Large ground finch



(b) Daphne Major

FIGURE 25.1 **A local population of the large ground finch.** **(a)** The large ground finch (*Geospiza magnirostris*) on Daphne Major. **(b)** Daphne Major, one of the Galápagos Islands.

Concept check: *What does the term local population mean?*

A Population Is a Group of Interbreeding Individuals That Share a Gene Pool

In genetics, the term *population* has a very specific meaning. With regard to sexually reproducing species, a **population** is a group of individuals of the same species that occupy the same region and can interbreed with one another. Many species occupy a wide geographic range and are divided into discrete populations. For example, distinct populations of a given species may be located on different continents, or populations on the same continent could be divided by a geographical feature such as a large mountain range.

A large population usually is composed of smaller groups called **local populations,** or **demes.** The members of a local population are far more likely to breed among themselves than with other members of the general population. Local populations are often separated from each other by moderate geographic barriers. The large ground finch (*Geospiza magnirostris*) is found on a small volcanic island called Daphne Major, which is one of the Galápagos Islands (Figure 25.1). Daphne Major is located northwest of the much larger Santa Cruz Island. The population of large ground finches on Daphne Major constitutes a local population of this species. Breeding is much more apt to occur among members of a local population than between members of neighboring populations. On relatively rare occasions, however, a bird may fly from Daphne Major to Santa Cruz Island, which would make breeding between the two different local populations possible.

Populations typically are dynamic units that change from one generation to the next. A population may change its size, geographic location, and genetic composition. With regard to size, natural populations commonly go through cycles of "feast or famine," during which environmental factors such as the amount of rainfall or food source cause the population to swell or shrink. In addition, natural predators or disease may periodically decrease the size of a population to significantly lower levels; the population later may rebound to its original size. Populations or individuals within populations may migrate to a new site and establish a distinct population in this location. The environment at this new geographic location may differ from the original site. What are the consequences of such changes? As population sizes and locations change, their genetic composition generally changes as well. As described later, population geneticists have developed mathematical models that predict how the gene pool will change in response to fluctuations in size, migration, and new environments.

At the Population Level, Some Genes May Be Monomorphic, But Most Are Polymorphic

In population genetics, the term **genetic polymorphism,** or simply **polymorphism** (meaning many forms), refers to the observation that many inherited traits display variation within a population. Historically, genetic polymorphism first referred to the variation in inherited traits that are observable with the naked eye. Polymorphisms in color and pattern have long attracted the attention of population geneticists. These include studies involving yellow and red varieties of the elder-flowered orchid, and brown, pink, and yellow land snails, which are discussed later in this chapter. Figure 25.2 illustrates a striking example of polymorphism in the Hawaiian happy-face spider (*Theridion grallator*). The three individuals shown in this figure are from the same species, but they differ in alleles that affect color and pattern.

What is the underlying cause of polymorphism? At the DNA level, polymorphism may be due to two or more alleles that influence the phenotype of the individual that inherits them. In other words, it is due to genetic variation. Geneticists also use the term **polymorphic** to describe a gene that commonly exists as two or more alleles in a population. By comparison, a **monomorphic** gene exists predominantly as a single allele in a population. By convention, when a single allele is found in at least 99% of all cases, the gene is considered monomorphic. (Some geneticists view an allele frequency of 95% or greater to be monomorphic.)

At the level of a particular gene, a polymorphism may involve various types of changes such as a deletion of a significant region of the gene, a duplication of a region, or a change in a single nucleotide. This last phenomenon is called a **single-nucleotide polymorphism (SNP).** SNPs are the smallest type of genetic change that can occur within a given gene and are also the most common. In humans, for example, SNPs represent 90% of all the variation in DNA sequences that occurs among different people. SNPs are found very frequently in genes. In the human population, a gene that is 2000 to 3000 bp in length, on average, contains 10 different sites that are polymorphic. The high frequency of SNPs indicates that polymorphism is the norm for most human genes. Likewise, relatively large, healthy populations of nearly all species exhibit a high level of genetic variation as evidenced by the occurrence of SNPs within most genes.

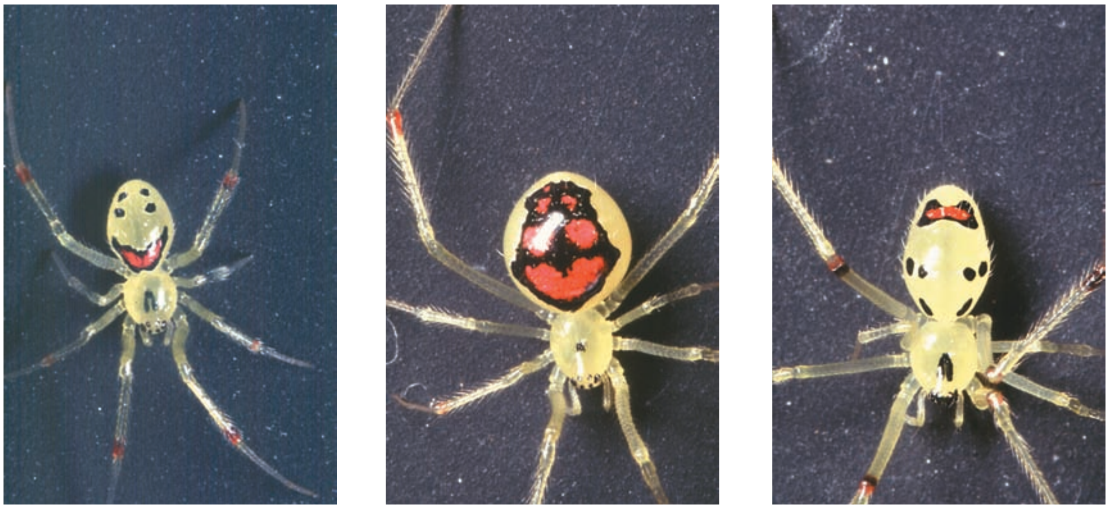


FIGURE 25.2 **Polymorphism in the Hawaiian happy-face spider.**

Genes→Traits These three spiders are members of the same species and carry the same genes. However, several genes that affect pigmentation patterns are polymorphic, meaning that more than one allele occurs in each gene within the population. This polymorphism within the Hawaiian happy-face spider population produces members that look quite different from each other.

Concept check: *Are polymorphisms common or rare in natural populations?*

Within a population, the alleles of a given gene may arise by different types of genetic changes. **Figure 25.3** considers a gene that exists in multiple forms in humans. This example is a short segment of DNA found within the human β-globin gene. The top sequence is an allele designated Hb^A, whereas the middle sequence is called Hb^S. These alleles differ from each other by a single nucleotide, so they are an example of a SNP. As discussed in Chapter 5, the Hb^S allele causes sickle cell disease in a homozygote. The bottom sequence contains a short, 5-bp deletion compared with the other two alleles. This deletion results in a nonfunctional β-globin polypeptide. Therefore, the bottom sequence is an example of a loss-of-function allele.

Population Genetics Is Concerned with Allele and Genotype Frequencies

As we have seen, population geneticists want to understand the prevalence of polymorphic genes within populations. Their goal is to identify the causative factors that govern changes in genetic variation. Much of their work evaluates the frequency of alleles in a quantitative way. Two fundamental calculations are central to population genetics: **allele frequencies** and **genotype frequencies.** The allele and genotype frequencies are defined as

$$\text{Allele frequency} = \frac{\text{Number of copies of an allele in a population}}{\text{Total number of all alleles for that gene in a population}}$$

$$\text{Genotype frequency} = \frac{\text{Number of individuals with a particular genotype in a population}}{\text{Total number of individuals in a population}}$$

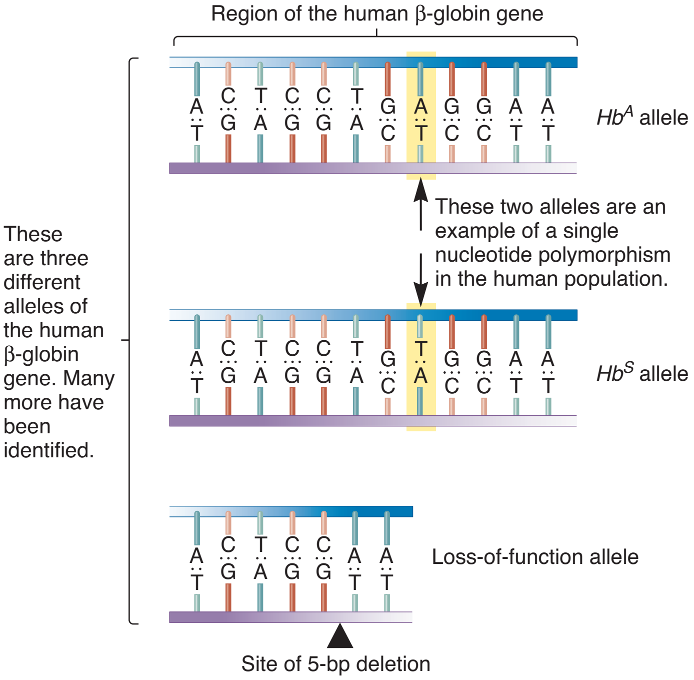

Region of the human β-globin gene
A C T C C T G A G G A A
T G A G G A C T C C T T
Hb^A allele
These two alleles are an example of a single nucleotide polymorphism in the human population.
These are three different alleles of the human β-globin gene. Many more have been identified.
A C T C C T G T G G A A
T G A G G A C A C C T T
Hb^S allele
A C T C C A A
T G A G G T T
Loss-of-function allele
Site of 5-bp deletion

FIGURE 25.3 **The relationship between alleles and various types of mutations.** The DNA sequence shown here is a small portion of the β-globin gene in humans. Mutations have altered the gene to create the three different alleles in this figure. The top two alleles differ by a single base pair and are referred to as a single-nucleotide polymorphism (SNP). The bottom allele has a 5-bp deletion that begins right after the arrowhead. The deletion results in a nonfunctional polypeptide. It is a loss-of-function allele.

Though these two frequencies are related, a clear distinction between them must be kept in mind. As an example, let's consider a hypothetical population of 100 frogs with the following genotypes:

64 dark green frogs with the genotype *GG*
32 medium green frogs with the genotype *Gg*
4 light green frogs with the genotype *gg*

When calculating an allele frequency, homozygous individuals have two copies of an allele, whereas heterozygotes have only one. For example, in tallying the *g* allele, each of the 32 heterozygotes has one copy of the *g* allele, and each light green frog has two copies. The allele frequency for *g* equals

$$g = \frac{32 + 2(4)}{2(64) + 2(32) + 2(4)}$$

$$= \frac{40}{200} = 0.2, \text{ or } 20\%$$

This result tells us that the allele frequency of *g* is 20%. In other words, 20% of the alleles for this gene in the population are the *g* allele.

Let's now calculate the genotype frequency of *gg* (light green) frogs:

$$gg = \frac{4}{64 + 32 + 4}$$

$$= \frac{4}{100} = 0.04, \text{ or } 4\%$$

We see that 4% of the individuals in this population are light green frogs.

Allele and genotype frequencies are always less than or equal to 1 (i.e., ≤100%). If a gene is monomorphic, the allele frequency for the single allele will equal or be close to a value of 1.0. For polymorphic genes, if we add up the frequencies for all of the alleles in the population, we should obtain a value of 1.0. In our frog example, the allele frequency of *g* equals 0.2. The frequency of the other allele, *G*, equals 0.8. If we add the two together, we obtain a value of $0.2 + 0.8 = 1.0$.

The Hardy-Weinberg Equation Can Be Used to Calculate Genotype Frequencies Based on Allele Frequencies

In 1908, a British mathematician, Godfrey Harold Hardy, and a German physician, Wilhelm Weinberg, independently derived a simple mathematical expression that predicts stability of allele and genotype frequencies from one generation to the next. It is called the **Hardy-Weinberg equilibrium,** because (under a given set of conditions, described later) the allele and genotype frequencies do not change over the course of many generations.

Why is the Hardy-Weinberg equilibrium a useful concept? An equilibrium is a null hypothesis, which suggests that evolutionary change is not occurring. In reality, however, populations rarely achieve an equilibrium. Therefore, the main usefulness of the Hardy-Weinberg equilibrium is that it provides a framework on which to understand changes in allele and genotype frequencies within a population when such an equilibrium is violated.

To appreciate the Hardy-Weinberg equilibrium, let's return to our hypothetical frog example in which a gene is polymorphic and exists as two different alleles: *G* and *g*. If the allele frequency of the allele *G* is denoted by the variable p, and the frequency of the allele *g* is denoted by q, then

$$p + q = 1$$

For example, if $p = 0.8$, then q must be 0.2. In other words, if the allele frequency of *G* equals 80%, the remaining 20% of alleles must be *g*, because together they equal 100%.

The **Hardy-Weinberg equation** is used to relate allele frequencies and genotype frequencies. For a diploid species, each individual inherits two copies of most genes. The Hardy-Weinberg equation assumes that the alleles for the next generation for any given individual are chosen randomly and independently of each other. Therefore, we can use the product rule (see Chapter 3) and multiply the sum, $p + q$, together. Because $p + q = 1$, we also know that their product also equals 1:

$$(p + q)(p + q) = 1$$

$$p^2 + 2pq + q^2 = 1 \text{ (Hardy-Weinberg equation)}$$

This equation applies to a gene in a diploid species that is found in only two alleles, which exist at frequencies designated p and q. As described in Solved Problem S4 at the end of the chapter, it can be expanded to three or more alleles.

The Hardy-Weinberg equation predicts an equilibrium—unchanging allele and genotype frequencies from generation to generation—if certain conditions are met in a population. With regard to the gene of interest, these conditions are as follows:

1. *No new mutations:* The gene of interest incurs no new mutations.
2. *No genetic drift:* The population is so large that allele frequencies do not change due to random sampling effects.
3. *No migration:* Individuals do not travel between different populations.
4. *No natural selection:* All of the genotypes have the same reproductive success.
5. *Random mating:* With respect to the gene of interest, the members of the population mate with each other randomly, without regard to their phenotypes and genotypes.

If the Hardy-Weinberg equation is related to our hypothetical frog population in which a gene exists in alleles designated *G* and *g*, then

p^2 equals the genotype frequency of *GG*
$2pq$ equals the genotype frequency of *Gg*
q^2 equals the genotype frequency of *gg*

If $p = 0.8$ and $q = 0.2$ and if the population is in Hardy-Weinberg equilibrium, then

$$GG = p^2 = (0.8)^2 = 0.64$$

$$Gg = 2pq = 2(0.8)(0.2) = 0.32$$

$$gg = q^2 = (0.2)^2 = 0.04$$

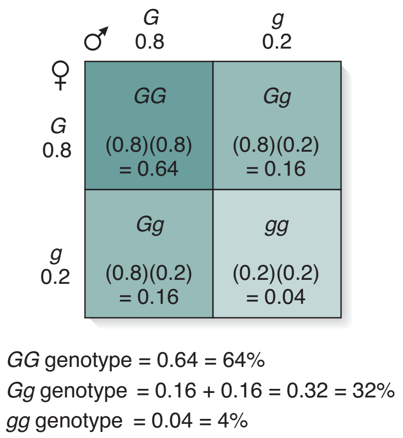


FIGURE 25.4 **A comparison between allele frequencies and the union of alleles in a Punnett square.** In a population in Hardy-Weinberg equilibrium, the frequency of gametes carrying a particular allele is equal to the allele frequency in the population.

In other words, if the allele frequency of *G* is 80% and the allele frequency of *g* is 20%, the genotype frequency of *GG* is 64%, *Gg* is 32%, and *gg* is 4%.

To illustrate the relationship between allele frequencies and genotypes, **Figure 25.4** compares the Hardy-Weinberg equation with the way that gametes combine randomly with each other to produce offspring. In a population, the frequency of a gamete carrying a particular allele is equal to the allele frequency in that population. In this example, the frequency of a gamete carrying the *G* allele equals 0.8.

We can use the product rule to determine the frequency of genotypes. For example, the frequency of producing a *GG* homozygote is $0.8 \times 0.8 = 0.64$, or 64%. Likewise, the probability of inheriting both *g* alleles is $0.2 \times 0.2 = 0.04$, or 4%. As seen in Figure 25.4, heterozygotes can be produced in two different ways. An offspring could inherit the *G* allele from its father and *g* from its mother, or *G* from its mother and *g* from its father. Therefore, the frequency of heterozygotes is $pq + pq$, which equals $2pq$; in our example, this is $2(0.8)(0.2) = 0.32$, or 32%.

The Hardy-Weinberg equation provides a quantitative relationship between allele and genotype frequencies in a population. **Figure 25.5** describes this relationship for different allele frequencies of *g* and *G*. As expected, when the allele frequency of *g* is very low, the *GG* genotype predominates; when the *g* allele frequency is high, the *gg* homozygote is most prevalent in the population. When the allele frequencies of *g* and *G* are intermediate in value, the heterozygote predominates.

In reality, no population satisfies the Hardy-Weinberg equilibrium completely. Nevertheless, in large natural populations with little migration and negligible natural selection, the Hardy-Weinberg equilibrium may be nearly approximated for certain genes.

The Hardy-Weinberg Equation Can Be Used to Detect Evolutionary Change and to Calculate the Frequencies of Heterozygous Carriers

In Chapter 3, the chi square test was used to determine if observed and expected data are in agreement. We can use a chi square test to determine whether a population exhibits Hardy-Weinberg equilibrium for a particular gene. If a population is not in equilibrium, this suggests evolutionary change. To carry out this calculation, it is necessary to distinguish between the homozygotes and heterozygotes, either phenotypically or at the molecular level. This is necessary so we can determine both the allele and genotype frequencies. As an example, let's consider a human blood type called the MN type. In this case, the blood type is determined by two codominant alleles, *M* and *N*. In an Inuit population in East Greenland, it was found that among 200 people, 168 were *MM*, 30 were *MN*, and 2 were *NN*. We can use these observed data to calculate the expected number of each genotype based on the Hardy-Weinberg equation.

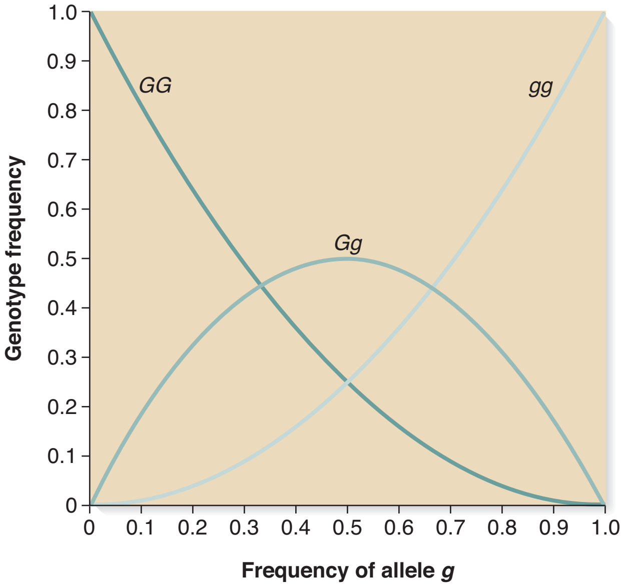


FIGURE 25.5 **The relationship between allele frequencies and genotype frequencies according to the Hardy-Weinberg equilibrium.** This graph assumes that *g* and *G* are the only two alleles for this gene.

$$\text{Allele frequency of } M = \frac{2(168) + 30}{400} = 0.915$$

$$\text{Allele frequency of } N = \frac{2(2) + 30}{400} = 0.085$$

Expected frequency of *MM* $= p^2 = (0.915)^2 = 0.837$

Expected number of *MM* individuals $= 0.837 \times 200 = 167.4$ (or 167 if rounded to the nearest individual)

Expected frequency of *NN* $= q^2 = (0.085)^2 = 0.007$

Expected number of *NN* individuals $= 0.007 \times 200 = 1.4$ (or 1 if rounded to the nearest individual)

Expected frequency of *MN* $= 2pq = 2(0.915)(0.085) = 0.156$

Expected number of *MN* individuals $= 0.156 \times 200 = 31.2$ (or 31 if rounded to the nearest individual)

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3}$$

$$\chi^2 = \frac{(168 - 167)^2}{167} + \frac{(30 - 31)^2}{31} + \frac{(2 - 1)^2}{1}$$

$$\chi^2 = 1.04$$

Let's refer back to Table 3.2 in Chapter 3 to evaluate the calculated chi square value. Because the gene exists in two alleles, the degrees of freedom equals 1. For a chi square of 1.04, the *P* value is between 0.5 and 0.2, which is well within the acceptable range. Therefore, we fail to reject the null hypothesis. In this case, the alleles for this gene appear to be in Hardy-Weinberg equilibrium.

When researchers have investigated other genes in various populations, a high chi square value may be obtained, and the hypothesis is rejected that the allele and genotype frequencies are in Hardy-Weinberg equilibrium. In these cases, we would say that such a population is in **disequilibrium.** Deviation from a Hardy-Weinberg equilibrium indicates evolutionary change. Factors such as natural selection, genetic drift, migration, and nonrandom mating may disrupt the Hardy-Weinberg equilibrium. Therefore, when population geneticists discover that a population is not in equilibrium, they try to determine which evolutionary factors are at work.

Another useful feature of the Hardy-Weinberg equation is that it allows us to estimate the frequency of heterozygotes for recessive genetic diseases, assuming a Hardy-Weinberg equilibrium. As an example, let's consider cystic fibrosis, which involves a gene that encodes a chloride transporter. Persons afflicted with this disorder have an irregularity in salt and water balance. One of the symptoms is thick mucus in the lungs that can contribute to repeated lung infections. In Caucasian populations, the frequency of affected individuals is approximately 1 in 2500. Because this is a recessive disorder, affected individuals are homozygotes. If q represents the allele frequency of the disease-causing allele, then

$$q^2 = 1/2500$$

$$q^2 = 0.0004$$

We take the square root to determine q:

$$q = \sqrt{0.0004}$$

$$q = 0.02$$

If p represents the normal allele,

$$p = 1 - q$$

$$p = 1 - 0.02 = 0.98$$

The frequency of heterozygous carriers is

$$2pq = 2(0.98)(0.02) = 0.0392, \text{ or } 3.92\%$$

25.1 REVIEWING THE KEY CONCEPTS

- All of the alleles of every gene in a population constitute the gene pool.
- For sexually reproducing organisms, a population is a group of individuals of the same species that occupy the same region and can interbreed with one another (see Figure 25.1).
- Genetic polymorphism refers to inherited traits or genes that exhibit variation in a population (see Figure 25.2).
- Single-nucleotide polymorphisms (SNPs) are the most common type of variation in genes (see Figure 25.3).
- Geneticists analyze genetic variation by determining allele and genotype frequencies.
- The Hardy-Weinberg equation can be used to calculate genotype frequencies based on allele frequencies (see Figures 25.4, 25.5).
- Deviation from Hardy-Weinberg equilibrium indicates that evolutionary change is occurring.
- The Hardy-Weinberg equation can be used to estimate the frequency of heterozygous carriers.

25.1 COMPREHENSION QUESTIONS

1. A gene pool is
   a. all of the genes in a single individual.
   b. all of the genes in the gametes from a single individual.
   c. all of the genes in a population of individuals.
   d. the random mixing of genes during sexual reproduction.
2. In natural populations, most genes are
   a. polymorphic.
   b. monomorphic.
   c. recessive.
   d. both a and c.
3. A gene exists in two alleles designated D and d. If 48 copies of this gene are the D allele and 152 are the d allele, what is the allele frequency of D?
   a. 0.24
   b. 0.32
   c. 0.38
   d. 0.76
4. The allele frequency of C is 0.4 and c is 0.6. If the population is in Hardy-Weinberg equilibrium, what is the frequency of heterozygotes?
   a. 0.16
   b. 0.24
   c. 0.26
   d. 0.48

25.2 OVERVIEW OF MICROEVOLUTION

The genetic variation in natural populations typically changes over the course of many generations. The term **microevolution** describes changes in a population's gene pool from generation to generation. Such change is rooted in two related phenomena (**Table 25.1**). First, the introduction of new genetic variation into a population is one essential aspect of microevolution. As discussed later in this chapter, gene variation can originate by a variety of molecular mechanisms. For example, new alleles of preexisting genes can arise by random mutations. Such events provide a continuous source of new variation to populations. However, due to their low rate of occurrence, mutations do not act as a major factor in promoting widespread changes in a population. If mutations were the only type of change occurring in a

TABLE **25.1**

Factors That Govern Microevolution

Source of New Allelic Variation*	
Mutation	Throughout most of this chapter, we consider allelic variation. Random mutations within preexisting genes introduce new alleles into populations, but at a very low rate. New mutations may be beneficial, neutral, or deleterious. For new alleles to rise to a significant percentage in a population, evolutionary mechanisms (i.e., natural selection, genetic drift, and/or migration) must operate on them.
Mechanisms That Alter Existing Genetic Variation	
Natural selection	This is the phenomenon in which the environment selects for individuals that possess certain traits that favor reproductive success. For example, natural selection may be related to the survival of members to reproductive age.
Genetic drift	This is a change in genetic variation from generation to generation due to random sampling error. Allele frequencies may change as a matter of chance from one generation to the next. This tends to have a greater effect in a small population.
Migration	Migration can occur between two different populations. The introduction of migrants into a recipient population may change the allele frequencies of that population.
Nonrandom mating	This is the phenomenon in which individuals select mates based on their phenotypes or genetic lineage. This can alter the relative proportion of homozygotes and heterozygotes predicted by the Hardy-Weinberg equation but does not change allele frequencies.

*Allelic variation is just one source of new genetic variation. Section 25.7 considers a variety of mechanisms through which new genetic variation can occur.

population, that population would not evolve at a significant rate because mutations are so rare.

Microevolution also involves the action of evolutionary mechanisms that alter the prevalence of a given allele or genotype in a population. These mechanisms are natural selection, genetic drift, migration, and nonrandom mating (see Table 25.1). The collective contributions of these evolutionary mechanisms over the course of many generations have the potential to promote widespread genetic changes in a population. In the following sections, we will examine how these mechanisms can affect the type of genetic variation that occurs when a gene exists in two or more alleles in a population. As you will learn, these mechanisms may cause a particular allele to be favored, or they may create a balance where two or more alleles are maintained in a population.

25.2 REVIEWING THE KEY CONCEPTS

- Mutations are the source of new genetic variation. However, the occurrence of new mutations does not greatly change allele frequencies because they are relatively rare. Other factors, such as natural selection, genetic drift, migration, and nonrandom mating, may alter allele and/or genotype frequencies (see Table 25.1).

25.2 COMPREHENSION QUESTIONS

1. Which of the following is a factor that does *not* promote widespread changes in allele or genotype frequencies?
   a. New mutation
   b. Natural selection
   c. Genetic drift
   d. Migration
   e. Nonrandom mating

25.3 NATURAL SELECTION

In the 1850s, Charles Darwin and Alfred Russel Wallace independently proposed the theory of evolution by **natural selection.** According to this theory, the conditions found in nature result in the selective survival and reproduction of individuals whose characteristics make them better adapted to their environment. These surviving individuals are more likely to reproduce and contribute offspring to the next generation. Natural selection can be related not only to differential survival but also to mating efficiency and fertility.

A modern restatement of the principles of natural selection can relate our knowledge of molecular genetics to the phenotypes of individuals.

1. Within a population, allelic variation arises in various ways, such as through random mutations that cause differences in DNA sequences. A mutation that creates a new allele may alter the amino acid sequence of the encoded protein, which, in turn, may alter the function of the protein.
2. Some alleles may encode proteins that enhance an individual's survival or reproductive capability compared with that of other members of the population. For example, an allele may produce a protein that is more efficient at a higher temperature, conferring on the individual a greater probability of survival in a hot climate.
3. Individuals with beneficial alleles are more likely to survive and contribute to the gene pool of the next generation.
4. Over the course of many generations, allele frequencies of many different genes may change through this process, thereby significantly altering the characteristics of a population or species. The net result of natural selection is a population that is better adapted to its environment and more successful at reproduction. Even so, it should be emphasized that species are not perfectly adapted to their environments, because mutations are random events and because the environment tends to change from generation to generation.

Fisher, Wright, and Haldane developed mathematical formulas to explain the theory of natural selection. As our knowledge of the process of natural selection has increased, it has become apparent that it operates in many different ways. In this section, we will consider a few examples of natural selection involving a single trait or a single gene that exists in two alleles. In reality, however, natural selection acts on populations of individuals in which many genes are polymorphic and each individual contains thousands or tens of thousands of different genes.

Darwinian Fitness Is a Measure of Reproductive Success

To begin our quantitative discussion of natural selection, we must examine the concept of **Darwinian fitness**—the relative likelihood that one genotype will contribute to the gene pool of the next generation compared with other genotypes. Natural selection acts on phenotypes that are derived from individuals' genotypes. Although Darwinian fitness often correlates with physical fitness, the two concepts are not identical. Darwinian fitness is a measure of reproductive success. An extremely fertile genotype may have a higher Darwinian fitness than a less fertile genotype that appears more physically fit.

To consider Darwinian fitness, let's use our example of a gene existing in the *A* and *a* alleles. If the three genotypes have the same level of mating success and fertility, we can assign fitness values to each genotype class based on their likelihood of surviving to reproductive age. For example, let's suppose that the relative survival to adulthood of each of the three genotype classes is as follows: For every five *AA* individuals that survive, four *Aa* individuals survive, and one *aa* individual survives. By convention, the genotype with the highest reproductive ability is given a fitness value of 1.0. Relative fitness values are denoted by the variable *W*. The fitness values of the other genotypes are assigned values relative to this 1.0 value:

$$\text{Fitness of } AA\text{: } W_{AA} = 5/5 = 1.0$$

$$\text{Fitness of } Aa\text{: } W_{Aa} = 4/5 = 0.8$$

$$\text{Fitness of } aa\text{: } W_{aa} = 1/5 = 0.2$$

Keep in mind that differences in reproductive success among genotypes may stem from various reasons. In this case, the fittest genotype is more likely to survive to reproductive age. In other situations, the most fit genotype is more likely to mate. For example, a bird with brightly colored feathers may have an easier time attracting a mate than a bird with duller plumage. Finally, a third possibility is that the fittest genotype may be more fertile, producing a higher number of gametes or gametes that are more successful at fertilization.

By studying species in their native environments, population geneticists have discovered that natural selection can occur in several ways. The patterns of natural selection depend on the relative fitness values of the different genotypes and on the variation of environmental effects. The four patterns of natural selection that we will consider are called directional, stabilizing, disruptive, and balancing selection. In most of the examples described next, natural selection leads to adaptation so that a species is better able to survive to reproductive age.

Directional Selection Favors the Extreme Phenotype

Directional selection favors individuals at one extreme of a phenotypic distribution that are more likely to survive and reproduce in a particular environment. Different phenomena may initiate the process of directional selection. One way that directional selection may arise is that a new allele may be introduced into a population by mutation, and the new allele may promote a higher fitness in individuals that carry it (**Figure 25.6**). If the homozygote carrying the favored allele has the highest fitness value, directional selection may cause this favored allele to eventually become the predominant allele in the population, perhaps eventually becoming a monomorphic allele.

Another possibility is that a population may be exposed to a prolonged change in its living environment. Under the new environmental conditions, the relative fitness values may change to favor one genotype, which will promote the elimination of other genotypes. As an example, let's suppose a population of finches on the mainland already has genetic variation in beak size. A small number of birds migrate to an island where the seeds are generally larger than they are on the mainland. In this new environment, birds with larger beaks have a higher fitness because they are better able to crack open the larger seeds and thereby survive to reproductive age. Over the course of many generations, directional selection produces a population of birds carrying alleles that promote larger beak size.

In the case of directional selection, allele frequencies may change in a step-by-step, generation-per-generation way. To appreciate how this occurs, let's take a look at how fitness affects the Hardy-Weinberg equilibrium and allele frequencies. Again, let's suppose a gene exists in two alleles: *A* and *a*. The three fitness values, which are based on relative survival levels, are

$$W_{AA} = 1.0$$

$$W_{Aa} = 0.8$$

$$W_{aa} = 0.2$$

In the next generation, we expect that the Hardy-Weinberg equilibrium is modified in the following way via directional selection:

Frequency of *AA*: p^2W_{AA}

Frequency of *Aa*: $2pqW_{Aa}$

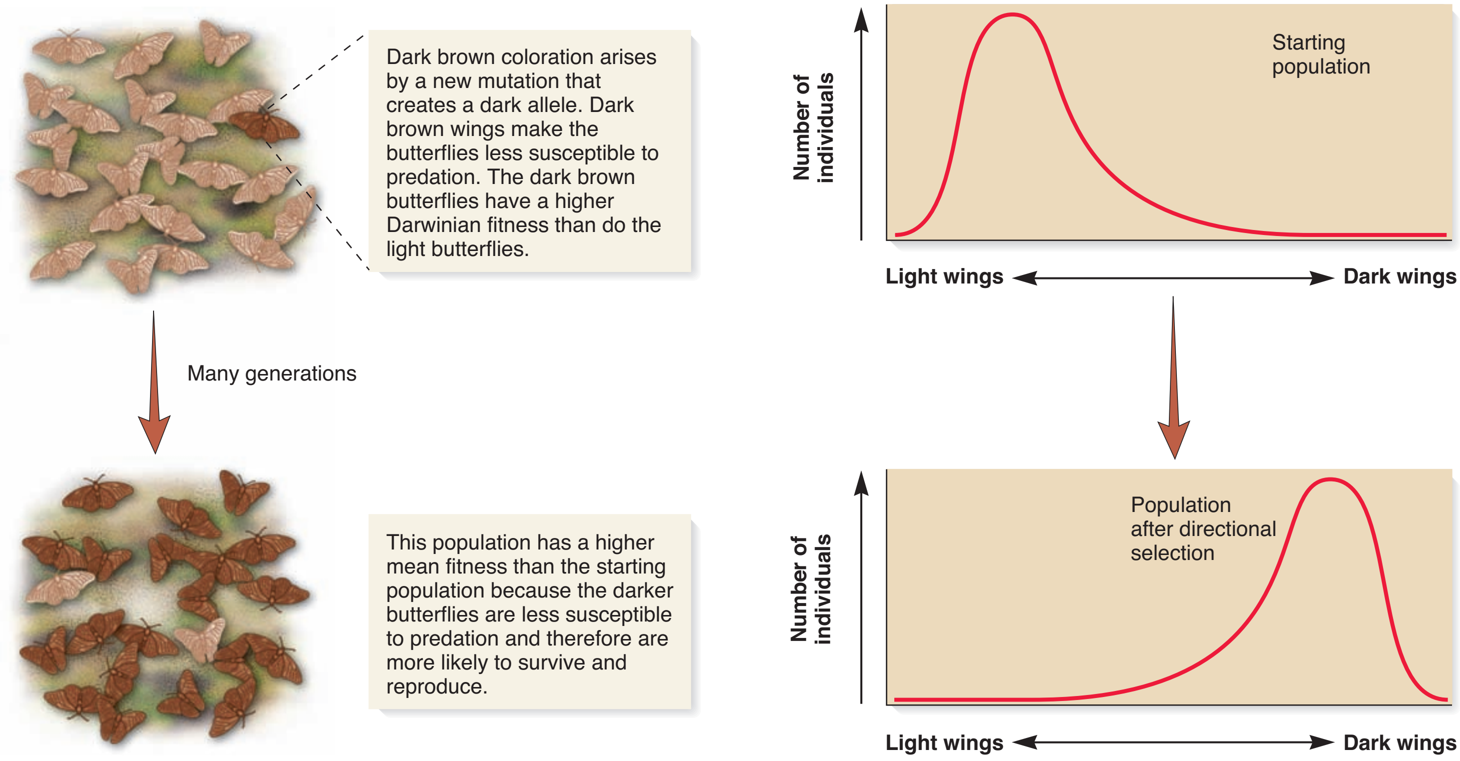
Frequency of *aa*: q^2W_{aa}

In a population that is changing due to natural selection, these three terms may not add up to 1.0, as they do in the Hardy-Weinberg equilibrium. Instead, the three terms sum to a value known as the **mean fitness of the population** ($\overline{W}$):

$$p^2W_{AA} + 2pqW_{Aa} + q^2W_{aa} = \overline{W}$$

Dividing both sides of the equation by the mean fitness of the population,

$$\frac{p^2W_{AA}}{\overline{W}} + \frac{2pqW_{Aa}}{\overline{W}} + \frac{q^2W_{aa}}{\overline{W}} = 1$$


Dark brown coloration arises by a new mutation that creates a dark allele. Dark brown wings make the butterflies less susceptible to predation. The dark brown butterflies have a higher Darwinian fitness than do the light butterflies.
Many generations
This population has a higher mean fitness than the starting population because the darker butterflies are less susceptible to predation and therefore are more likely to survive and reproduce.
Number of individuals
Starting population
Light wings
Dark wings
Number of individuals
Population after directional selection
Light wings
Dark wings

(a) An example of directional selection

(b) Graphical representation of directional selection

FIGURE 25.6 **Directional selection.** **(a)** A new mutation arises in a population that confers higher Darwinian fitness. In this example, butterflies with dark wings are more likely to survive and reproduce. Over many generations, directional selection favors the prevalence of darker individuals. **(b)** A graphical representation of directional selection.

Concept check: *In this form of natural selection, explain the meaning of the word directional.*

Using this equation, we can calculate the expected genotype and allele frequencies after one generation of directional selection:

$$\text{Frequency of } AA \text{ genotype: } \frac{p^2 W_{AA}}{\overline{W}}$$

$$\text{Frequency of } Aa \text{ genotype: } \frac{2pq W_{Aa}}{\overline{W}}$$

$$\text{Frequency of } aa \text{ genotype: } \frac{q^2 W_{aa}}{\overline{W}}$$

$$\text{Allele frequency of } A\text{: } \quad p_A = \frac{p^2 W_{AA}}{\overline{W}} + \frac{pq W_{Aa}}{\overline{W}}$$

$$\text{Allele frequency of } a\text{: } \quad q_a = \frac{q^2 W_{aa}}{\overline{W}} + \frac{pq W_{Aa}}{\overline{W}}$$

As an example, let's suppose that the starting allele frequencies are $A = 0.5$ and $a = 0.5$, and use fitness values of 1.0, 0.8, and 0.2 for the three genotypes, *AA*, *Aa*, and *aa*, respectively. We begin by calculating the mean fitness of the population:

$$p^2 W_{AA} + 2pq W_{Aa} + q^2 W_{aa} = \overline{W}$$

$$\overline{W} = (0.5)^2(1) + 2(0.5)(0.5)(0.8) + (0.5)^2(0.2)$$

$$\overline{W} = 0.25 + 0.4 + 0.05 = 0.7$$

After one generation of directional selection,

$$\text{Frequency of } AA \text{ genotype: } \frac{p^2 W_{AA}}{\overline{W}} = \frac{(0.5)^2(1)}{0.7} = 0.36$$

$$\text{Frequency of } Aa \text{ genotype: } \frac{2pq W_{Aa}}{\overline{W}} = \frac{2(0.5)(0.5)(0.8)}{0.7} = 0.57$$

$$\text{Frequency of } aa \text{ genotype: } \frac{q^2 W_{aa}}{\overline{W}} = \frac{(0.5)^2(0.2)}{0.7} = 0.07$$

$$\text{Allele frequency of } A\text{: } \quad p_A = \frac{p^2 W_{AA}}{\overline{W}} + \frac{pq W_{Aa}}{\overline{W}}$$

$$= \frac{(0.5)^2(1)}{0.7} + \frac{(0.5)(0.5)(0.8)}{0.7} = 0.64$$

$$\text{Allele frequency of } a\text{: } \quad q_a = \frac{q^2 W_{aa}}{\overline{W}} + \frac{pq W_{Aa}}{\overline{W}}$$

$$= \frac{(0.5)^2(0.2)}{0.7} + \frac{(0.5)(0.5)(0.8)}{0.7} = 0.36$$

After one generation, the allele frequency of *A* has increased from 0.5 to 0.64, and the frequency of *a* has decreased from 0.5 to 0.36. This has occurred because the *AA* genotype has the highest fitness, whereas the *Aa* and *aa* genotypes have lower fitness values. Another interesting feature of natural selection is that it raises the mean fitness of the population. If we assume the individual fitness values are constant, the mean fitness of this next generation is

$$\begin{aligned}\overline{W} &= p^2 W_{AA} + 2pqW_{Aa} + q^2 W_{aa}\\ &= (0.64)^2(1) + 2(0.64)(0.36)(0.8) + (0.36)^2(0.2)\\ &= 0.80\end{aligned}$$

The mean fitness of the population has increased from 0.7 to 0.8.

What are the consequences of directional selection at the population level? This population is better adapted to its environment than the previous one. Another way of viewing this calculation is that the subsequent population has a greater reproductive potential than the previous one. We could perform the same types of calculations to find the allele frequencies and mean fitness value in the next generation. If we assume the individual fitness values remain constant, the frequencies of *A* and *a* in the next generation are 0.85 and 0.15, respectively, and the mean fitness increases to 0.931. As we can see, the general trend is to increase *A*, decrease *a*, and increase the mean fitness of the population.

In the previous example, we considered the effects of directional selection by beginning with allele frequencies at intermediate levels (namely, $A = 0.5$ and $a = 0.5$). **Figure 25.7** illustrates what would happen if a new mutation introduced the *A* allele into a population that was originally monomorphic for the *a* allele. As before, the *AA* homozygote has a fitness of 1.0, the *Aa* heterozygote 0.8, and the recessive *aa* homozygote 0.2. Initially, the *A* allele is at a very low frequency in the population. If it is not lost initially due to genetic drift, its frequency slowly begins to rise and then, at intermediate values, rises much more rapidly.

Eventually, this type of natural selection may lead to fixation of a beneficial allele. Fixation occurs when only one of the alleles remains; its frequency is fixed at 100%. However, a new beneficial allele is in a precarious situation when its frequency is very low. As discussed later, genetic drift is likely to eliminate new mutations, even beneficial ones, due to chance fluctuations.

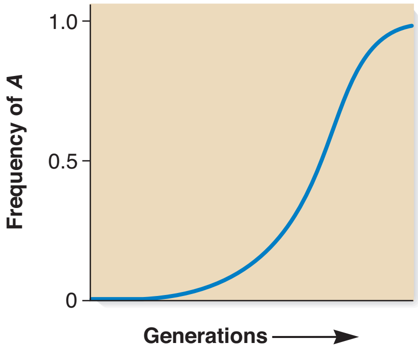


FIGURE 25.7 The fate of a beneficial allele that is introduced as a new mutation into a population. A new allele (*A*) is beneficial in the homozygous condition: $W_{AA} = 1.0$. The heterozygote, *Aa* ($W_{Aa} = 0.8$), and homozygote, *aa* ($W_{aa} = 0.2$), have lower fitness values.

Researchers have identified many examples of directional selection in nature. Resistance to antibiotics is a growing concern in the treatment of infection. The selection of bacterial strains resistant to one or more antibiotics typically occurs in a directional manner. Similarly, the resistance of insects to pesticides, such as DDT (dichlorodiphenyltrichloroethane), occurs in a directional manner. DDT usage began in the 1940s as a way to decrease the populations of mosquitoes and other insects. However, certain insect species can become resistant to DDT by a dominant mutation in a single enzyme-encoding gene. The mutant enzyme detoxifies DDT, making it harmless to the insect. **Figure 25.8** shows the results of an experiment in which mosquito larvae (*Aedes aegypti*) were exposed to DDT over the course of seven generations. The starting population showed a low level of DDT resistance, as evidenced by the low percentage of survivors after exposure to DDT. By comparison, in seven generations, nearly 100% of the population was DDT-resistant. These results illustrate the power of directional selection in promoting change in a population. Since the 1950s, resistance to nearly every known insecticide has evolved within 10 years of its commercial introduction!

Stabilizing Selection Favors Individuals with Intermediate Phenotypes

In **stabilizing selection,** the extreme phenotypes for a trait are selected against, and those individuals with the intermediate

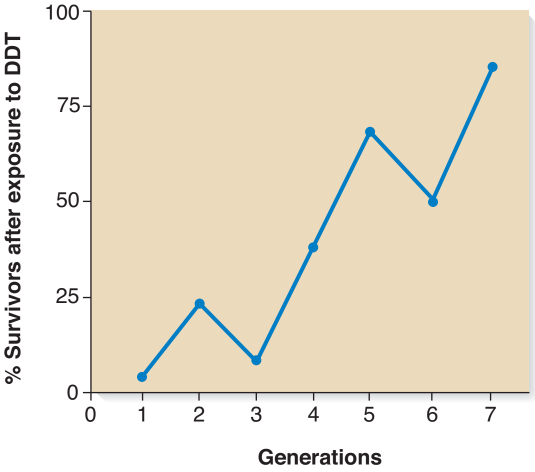


FIGURE 25.8 Directional selection for DDT resistance in a mosquito population. In this experiment, mosquito larvae (*Aedes aegypti*) were exposed to 10 mg/L of DDT. The percentage of survivors was recorded, and then the survivors of each generation were used as parents for the next generation.

Concept check: In this example, is directional selection promoting genetic diversity? Explain.

phenotypes have the highest fitness values. Stabilizing selection tends to decrease genetic diversity for a particular gene because it eliminates alleles that cause a greater variation in phenotypes. An example of stabilizing selection involves clutch size in birds, which was first proposed by British biologist David Lack in 1947. Under stabilizing selection, birds that lay too many or too few eggs have lower fitness values than those that lay an intermediate number (**Figure 25.9**). Laying too many eggs may cause many offspring to die due to inadequate parental care and food. In addition, the strain on the parents themselves may decrease their likelihood of survival and therefore their ability to produce more offspring. Having too few offspring, on the other hand, does not contribute many individuals to the next generation. Therefore, the most successful parents are those that produce an intermediate clutch size. In the 1980s, Swedish evolutionary biologist Lars Gustafsson and colleagues examined the phenomenon of stabilizing selection in the collared flycatcher, *Ficedula albicollis*, on the island of Gotland, which is southeast of the mainland of Sweden. They discovered that Lack's hypothesis that clutch size is subject to the action of stabilizing selection appears to be true for this species.

Disruptive Selection Favors Multiple Phenotypes

Disruptive selection, also known as diversifying selection, favors the survival of two or more different genotypes that produce different phenotypes (**Figure 25.10**). In disruptive selection, the fitness values of a particular genotype are higher in one environment and lower in a different one. Disruptive selection is likely to occur in populations that occupy diverse environments so that some members of the species survive in each type of environmental condition.

As an example, **Figure 25.11a** shows a photograph of land snails, *Cepaea nemoralis*, that live in woods and open fields. This snail is polymorphic in color and banding patterns. In 1954, Arthur Cain and Philip Sheppard found that snail color was correlated with the environment. As shown in **Figure 25.11b**, the highest frequency of brown shell color was found in snails in the beechwoods, where there are wide expanses of dark soil. Their frequency was substantially less in other environments. By comparison, pink snails are most common in the leaf litter of forest floors, and the yellow snails are most abundant in the sunny, grassy areas of hedgerows and rough herbage. Researchers have suggested that this disruptive selection can be explained by different levels of predation by thrushes. Depending on the environment, certain snail phenotypes may be more easily seen by their predators than others. Migration can occasionally occur between the snail populations, which maintains the polymorphism among these different environments.

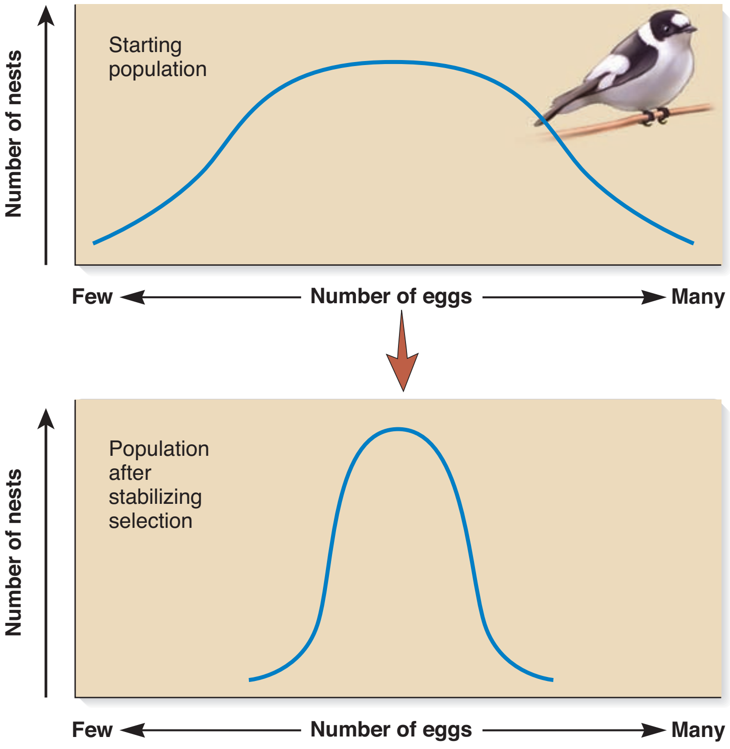


FIGURE 25.9 **Stabilizing selection.** In this pattern of natural selection, the extremes of a phenotypic distribution are selected against. Those individuals with intermediate traits have the highest fitness. This results in a population with less diversity and more uniform traits.

Concept check: In general, why does stabilizing selection decrease genetic diversity?

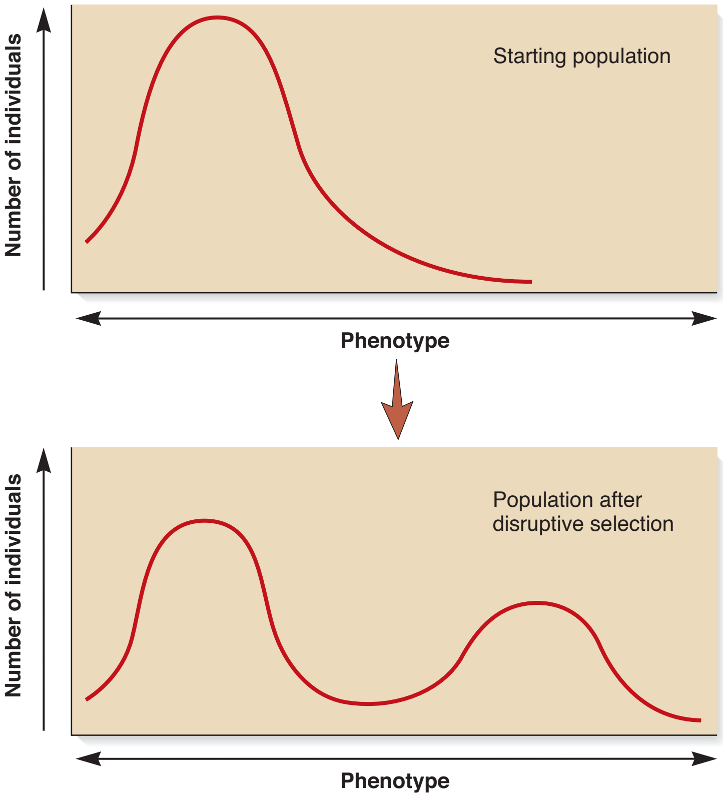
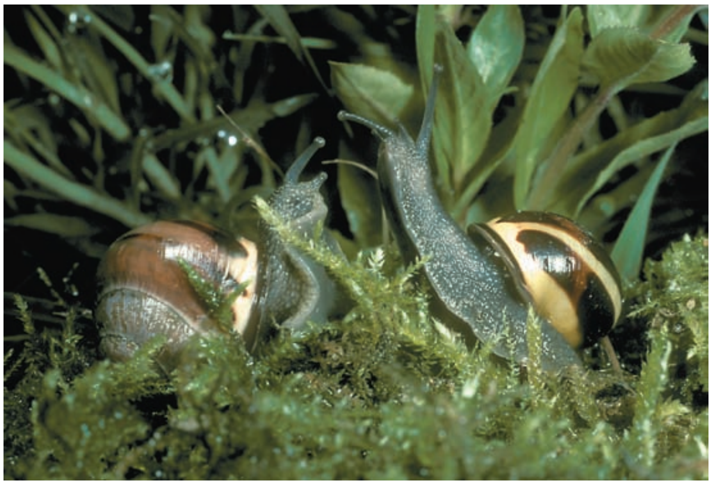


FIGURE 25.10 **Disruptive selection.** Over time, this form of selection favors two or more phenotypes due to heterogeneous environments.

Concept check: Does this form of natural selection favor polymorphism? Explain why or why not.



(a) Land snails

Habitat	*Brown*	*Pink*	*Yellow*
Beechwoods	0.23	0.61	0.16
Deciduous woods	0.05	0.68	0.27
Hedgegrows	0.05	0.31	0.64
Rough herbage	0.004	0.22	0.78

(b) Frequency of snail color

FIGURE 25.11 **Polymorphism in the land snail, *Cepaea nemoralis.*** **(a)** This species of snail can exist in several different colors and banding patterns. **(b)** Coloration of the snails is correlated with the specific environments where they are located.

Genes→Traits Snail coloration is an example of genetic polymorphism due to heterogeneous environments; the genes governing shell coloration are polymorphic. The predation of snails is correlated with their ability to be camouflaged in their natural environment. Snails with brown shells are most prevalent in beechwoods, where the soil is dark. Pink snails are most abundant in the leaf litter of beechwoods and deciduous woods. Yellow snails are most prevalent in sunnier locations, such as hedgerows and rough herbage.

Balancing Selection May Occur Due to Heterozygote Advantage or Negative Frequency-Dependent Selection

As we have just seen, genetic polymorphisms may occur when a species occupies diverse environments. Researchers have discovered other patterns of natural selection that favor the maintenance of two or more alleles in a more homogeneous environment. These patterns, collectively called **balancing selection,** also result in a genetic polymorphism in a population.

For genetic variation involving a single gene, balancing selection may arise when the heterozygote has a higher fitness than either corresponding homozygote, a situation called **heterozygote advantage** or **overdominance.** In this case, an equilibrium is reached in which both alleles are maintained in the population. If the fitness values are known for each of the genotypes, the allele frequencies at equilibrium can be calculated. To do so, we must consider the **selection coefficient** (***s***)**,** which measures the degree to which a genotype is selected against.

$$s = 1 - W$$

By convention, the genotype with the highest fitness has an s value of zero. Genotypes at a selective disadvantage have s values that are greater than 0 but less than or equal to 1.0. An extreme case is a recessive lethal allele. It would have an s value of 1.0 in the homozygote, whereas the s value in the heterozygote could be 0.

Let's consider the following case of relative fitness, where

$$W_{AA} = 0.7$$
$$W_{Aa} = 1.0$$
$$W_{aa} = 0.4$$

The selection coefficients are

$$s_{AA} = 1 - 0.7 = 0.3$$
$$s_{Aa} = 1 - 1.0 = 0$$
$$s_{aa} = 1 - 0.4 = 0.6$$

The population reaches an equilibrium when

$$s_{AA}p = s_{aa}q$$

If we take this equation, let $q = 1 - p$, and then solve for p:

$$p = \text{Allele frequency of } A = \frac{s_{aa}}{s_{AA} + s_{aa}}$$
$$= \frac{0.6}{0.3 + 0.6} = 0.67$$

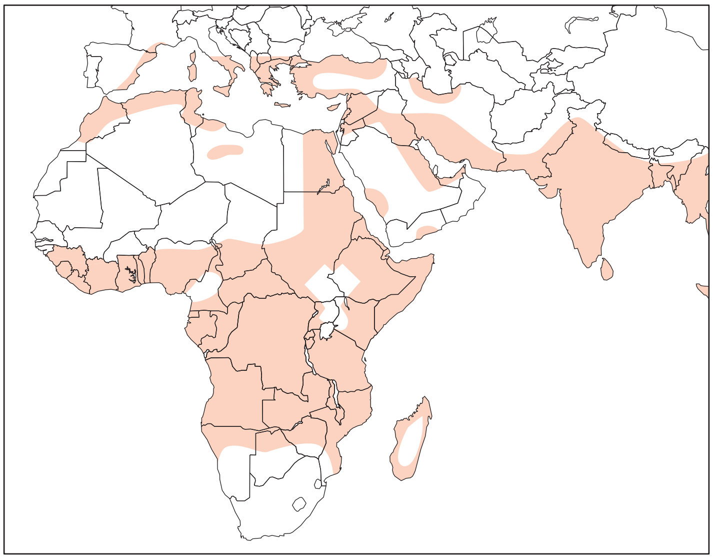
If we let $p = 1 - q$ and then solve for q:

$$q = \text{Allele frequency of } a = \frac{s_{AA}}{s_{AA} + s_{aa}}$$
$$= \frac{0.3}{0.3 + 0.6} = 0.33$$

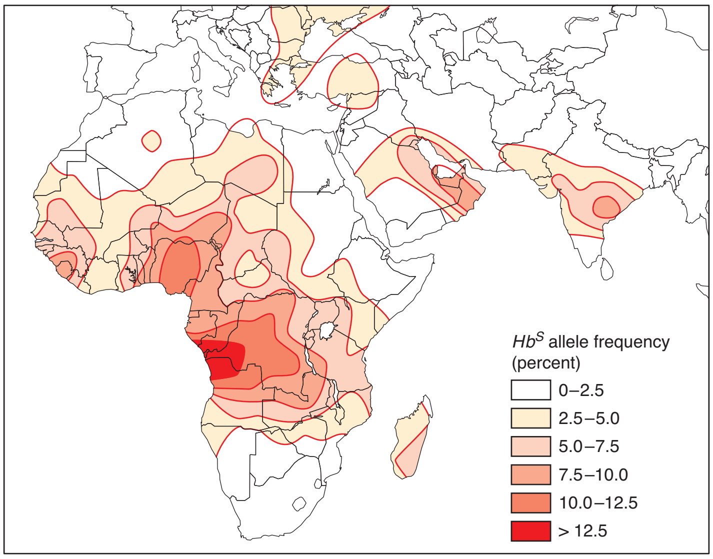
In this example, balancing selection maintains the two alleles in the population at frequencies in which A equals 0.67 and a equals 0.33.

Heterozygote advantage can sometimes explain the high frequency of alleles that are deleterious in a homozygous condition. A classic example is the Hb^S allele of the human β-globin gene. A homozygous Hb^SHb^S individual displays sickle cell disease, a disorder that leads to the sickling of the red blood cells. The Hb^SHb^S homozygote has a lower fitness than a homozygote with two copies of the more common β-globin allele, Hb^AHb^A. However, the heterozygote, Hb^AHb^S, has a higher level of fitness than either homozygote in areas where malaria is endemic (**Figure 25.12**). Compared with Hb^AHb^A homozygotes, heterozygotes have a 10% to 15% better chance of survival if infected by the malarial parasite, *Plasmodium falciparum*. Therefore, the Hb^S allele is maintained in populations where malaria is prevalent, even though the allele is detrimental in the homozygous state.

In addition to sickle cell disease, other gene mutations that cause human disease in the homozygous state are thought to be prevalent because of heterozygote advantage. For example,



(a) Malaria prevalence



(b) Hb^S allele frequency

FIGURE 25.12 **The geographic relationship between malaria and the frequency of the sickle cell allele in human populations.** **(a)** The geographic prevalence of malaria in Africa and surrounding areas. **(b)** The frequency of the Hb^S allele in the same areas.

Genes→Traits The sickle cell allele of the β-globin gene is maintained in human populations by balancing selection. In areas where malaria is prevalent, the heterozygote carrying one copy of the Hb^S allele has a greater fitness than either of the corresponding homozygotes (Hb^AHb^A and Hb^SHb^S). Therefore, even though the Hb^SHb^S homozygotes suffer the detrimental consequences of sickle cell disease, this negative aspect is balanced by the beneficial effects of malarial resistance in the heterozygotes.

Concept check: *Explain why the Hb^S allele is prevalent in certain regions even though it is detrimental in the homozygous condition.*

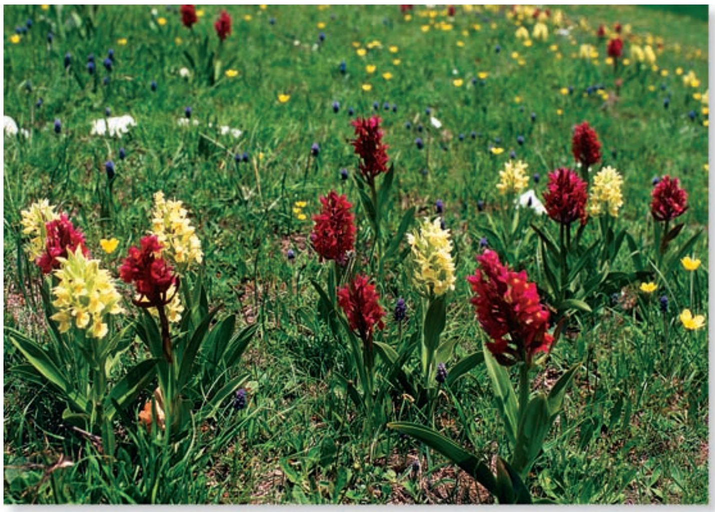


FIGURE 25.13 **The two color variations found in the elder-flowered orchid, *Dactylorhiza sambucina.*** The two colors are maintained in the population due to negative frequency-dependent selection.

Concept check: *Explain how negative frequency-dependent selection works.*

the high prevalence of the allele causing cystic fibrosis may be related to this phenomenon, but the advantage that a heterozygote may possess is not understood.

Negative frequency-dependent selection is a second mechanism of balancing selection. In this pattern of natural selection, the fitness of a genotype decreases when its frequency becomes higher. In other words, rare individuals have a higher fitness than more common individuals. Therefore, rare individuals are more likely to reproduce, whereas common individuals are less likely, thereby producing a balanced polymorphism in which no genotype becomes too rare or too common.

An interesting example of negative frequency-dependent selection involves the elder-flowered orchid, *Dactylorhiza sambucina* (Figure 25.13). Throughout its range, both yellow- and red-flowered individuals are prevalent. The explanation for this polymorphism is related to its pollinators, which are mainly bumblebees such as *Bombus lapidarius* and *B. terrestris*. The pollinators increase their visits to the flower color of *D. sambucina* as it becomes less common in a given area. One reason why this may occur is because *D. sambucina* is a rewardless flower; that is, it does not provide its pollinators with any reward for visiting, such as sweet nectar. Pollinators learn that the more common color of *D. sambucina* in a given area does not offer a reward, and they increase their visits to the less-common flower. For this reason, the relative fitness of the less-common flower increases.

The Grants Have Observed Natural Selection in Galápagos Finches

Let's now turn to a study that demonstrated natural selection in action. Since 1973, Peter Grant, Rosemary Grant, and their colleagues have studied the process of natural selection in finches

FIGURE 25.14 The medium ground finch (*Geospiza fortis*), which is found on Daphne Major.

found on the Galápagos Islands. For over 30 years, the Grants have focused much of their research on one of the Galápagos Islands known as Daphne Major. This small island (0.34 km^2) has a moderate degree of isolation (8 km from the nearest island), an undisturbed habitat, and a resident population of finches, including the medium ground finch, *Geospiza fortis* (**Figure 25.14**).

To study natural selection, the Grants observed various traits in the medium ground finch, including beak size, over the course of many years. The medium ground finch has a relatively small crushing beak, suitable for breaking open small, tender seeds. The Grants quantified beak size among the medium ground finches of Daphne Major by carefully measuring beak depth (a measurement of the beak from top to bottom, at its base) on individual birds. During the course of their studies, they compared the beak sizes of parents and offspring by examining many broods over several years. The depth of the beak was transmitted from parents to offspring, regardless of environmental conditions, indicating that differences in beak sizes are due to genetic differences in the population. In other words, they found that beak depth is a heritable trait.

By measuring many birds every year, the Grants were able to assemble a detailed portrait of natural selection from generation to generation. In the study shown in **Figure 25.15**, they measured beak depth in 1976 and 1978.

▶ THE HYPOTHESIS

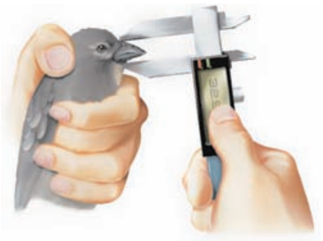
Beak size will be influenced by natural selection. Environments that produce larger seeds will select for birds with large beaks.

▶ TESTING THE HYPOTHESIS — FIGURE 25.15 Natural selection in medium ground finches of Daphne Major.

Experimental level **Conceptual level**

1. In 1976, measure beak depth in parents and offspring of the species *G. fortis*.

Capture birds and measure beak depth.



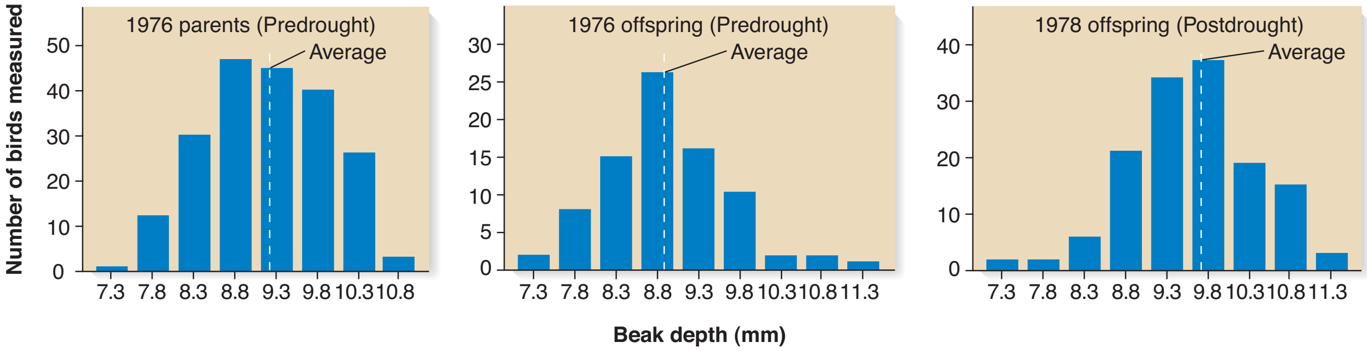
This is a way to measure a trait that may be subject to natural selection.

2. Repeat the procedure on offspring that were born in 1978 and had reached mature size. A drought had occurred in 1977. The plants that produced larger seeds tolerated the drought better and were more productive.

Capture birds and measure beak depth.

This is a way to measure a trait that may be subject to natural selection.

▶ THE DATA



Data from B. Rosemary Grant and Peter R. Grant (2003) What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. *Bioscience 53,* 965–975.

▶ INTERPRETING THE DATA

In the wet year of 1976, the plants of Daphne Major produced the small seeds that these finches were able to eat in abundance. However, a drought occurred in 1977. During this year, the plants on Daphne Major tended to produce few of the smaller seeds, which the finches rapidly consumed. To survive, the finches resorted to eating larger, drier seeds, which were harder to crush. As a result, the birds that survived tended to have larger beaks, because they were better able to break open these large seeds. In the year after the drought, the average beak depth of offspring in the population increased to approximately 9.8 mm because the surviving birds with larger beaks passed this trait on to their offspring. This is likely to be due to directional selection (see Figure 25.6), although genetic drift, which is discussed next, could also contribute to these data. Overall, these results illustrate the power of natural selection to alter the nature of a trait, in this case, beak depth, in a given population.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

25.3 REVIEWING THE KEY CONCEPTS

- Natural selection is a process that changes allele frequencies from one generation to the next based on fitness, which is the relative reproductive successes of different genotypes.
- Directional selection favors the extreme phenotype (see Figures 25.6–25.8).
- Stabilizing selection favors individuals with intermediate phenotypes (see Figure 25.9).
- Disruptive selection favors multiple phenotypes. This may occur due to heterogeneous environments (see Figures 25.10, 25.11).
- Balancing selection results in a stable polymorphism. Examples include heterozygote advantage and negative-frequency dependent selection (see Figures 25.12, 25.13).
- The Grants observed natural selection in a finch population. The selection involved a change in beak size due to drought conditions (see Figures 25.14, 25.15).

25.3 COMPREHENSION QUESTIONS

1. Darwinian fitness is a measure of
   a. survival.
   b. reproductive success.
   c. heterozygosity of the gene pool.
   d. polymorphisms in a population.
2. Within a particular population, darkly colored rats are more likely to survive than more lightly colored individuals. This scenario is likely to result in
   a. directional selection.
   b. stabilizing selection.
   c. disruptive selection.
   d. balancing selection.
3. A population occupies a diverse environment in which the fitness of some genotypes is higher in one environment while the fitness of other individuals is higher in another environment. This scenario is likely to result in
   a. directional selection.
   b. stabilizing selection.
   c. disruptive selection.
   d. balancing selection.
4. A gene exists in two alleles and the heterozygote has the highest fitness. This scenario is likely to result in
   a. directional selection.
   b. stabilizing selection.
   c. disruptive selection.
   d. balancing selection.

25.4 GENETIC DRIFT

In the 1930s, geneticist Sewall Wright played a key role in developing the concept of **random genetic drift,** or simply, **genetic drift,** which refers to changes in allele frequencies in a population due to random fluctuations. As a matter of chance, the frequencies of alleles found in gametes that unite to form zygotes vary from generation to generation. Over the long run, genetic drift usually results in either the loss of an allele or its fixation at 100% in the population. The process is random with regard to particular alleles. Genetic drift can lead to the loss or fixation of deleterious, neutral, or beneficial alleles. The rate at which this occurs depends on the population size and on the initial allele frequencies. **Figure 25.16** illustrates the potential consequences of genetic drift in one large ($N = 1000$) and five small ($N = 20$) populations. At the beginning of this hypothetical simulation, all of these populations have identical allele frequencies: $A = 0.5$ and $a = 0.5$. In the five small populations, this allele frequency fluctuates substantially from generation to generation. Eventually, one of the alleles is eliminated and the other is fixed at 100%. At this point, the allele has become monomorphic and cannot fluctuate any further. By comparison, the allele frequencies in the large population fluctuate much less, because random sampling error is expected to have a smaller effect. Nevertheless, genetic drift leads to homozygosity even in large populations, but this takes many more generations to occur.

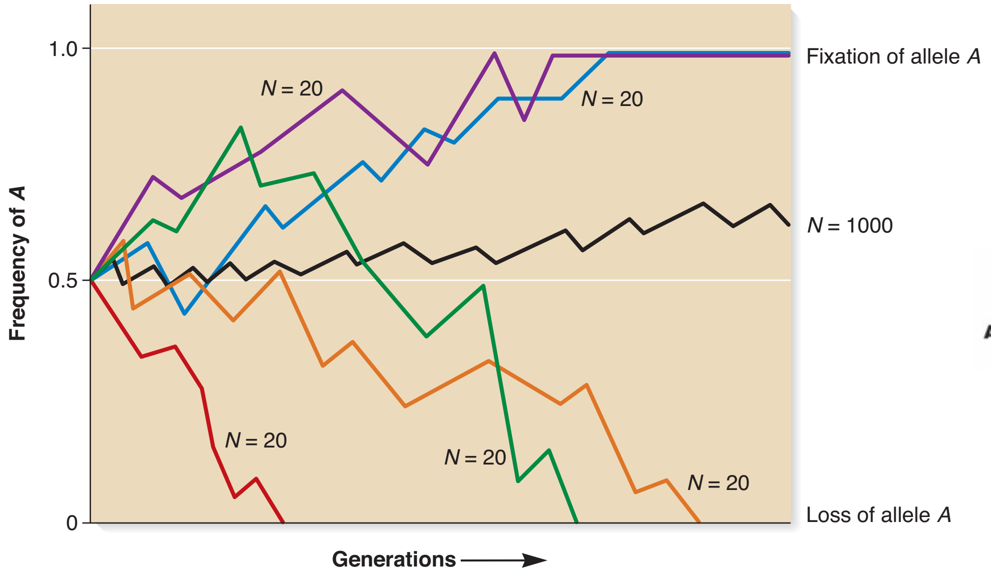
Now let's ask two questions:

1. How many new mutations do we expect in a natural population?
2. How likely is it that any new mutation will be either fixed in, or eliminated from, a population due to genetic drift?

With regard to the first question, the average number of new mutations depends on the mutation rate (μ) and the number of individuals in a population (N). If each individual has two copies of the gene of interest, the expected number of new mutations in this gene is

$$\text{Expected number of new mutations} = 2N\mu$$

From this, we see that a new mutation is more likely to occur in a large population than in a small one. This makes sense, because the larger population has more copies of the gene to be mutated.



ONLINE ANIMATION

FIGURE 25.16 **A hypothetical simulation of genetic drift.** In all cases, the starting allele frequencies are $A = 0.5$ and $a = 0.5$. The colored lines illustrate five populations in which $N = 20$; the black line shows a population in which $N = 1000$.

Concept check: How does population size affect genetic drift?

With regard to the second question, the probability of fixation of a newly arising allele due to genetic drift is

$$\text{Probability of fixation} = \frac{1}{2N} \text{ (assuming equal numbers of males and females contribute to the next generation)}$$

In other words, the probability of fixation is the same as the initial allele frequency in the population. For example, if $N = 20$, the probability of fixation of a new allele equals $1/(2 \times 20)$, or 2.5%. Conversely, a new allele may be lost from the population.

$$\text{Probability of elimination} = 1 - \text{probability of fixation}$$
$$= 1 - \frac{1}{2N}$$

If $N = 20$, the probability of elimination equals $1 - 1/(2 \times 20)$, or 97.5%. As you may have noticed, the value of N has opposing effects with regard to new mutations and their eventual fixation in a population. When N is very large, new mutations are much more likely to occur. Each new mutation, however, has a greater chance of being eliminated from the population due to genetic drift. On the other hand, when N is small, the probability of new mutations is also small, but if they occur, the likelihood of fixation is relatively large.

Now that we appreciate the phenomenon of genetic drift, we can ask a third question:

3. If fixation of a new allele does occur, how many generations is it likely to take?

The formula for calculating this also depends on the number of individuals in the population:

$$\bar{t} = 4N$$

where

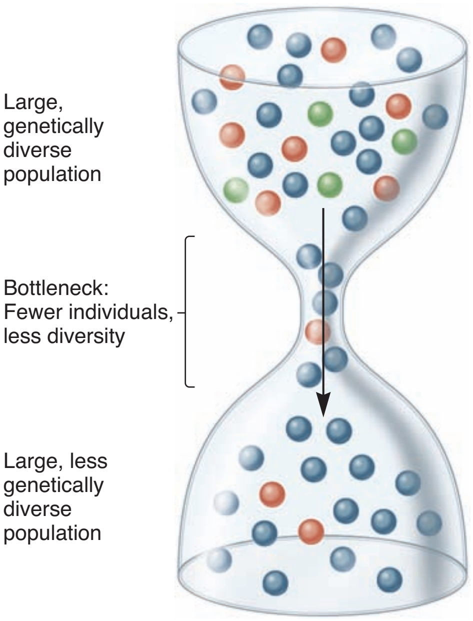
$\bar{t}$ equals the average number of generations to achieve fixation

N equals the number of individuals in the population, assuming that males and females contribute equally to each succeeding generation

As you may have expected, allele fixation takes much longer in large populations. If a population has 1 million breeding members, it takes, on average, 4 million generations, perhaps an insurmountable period of time, to reach fixation. In a small group of 100 individuals, however, fixation takes only 400 generations, on average.

In nature, geography and population size can influence how genetic drift affects the genetic composition of a species in different ways. Some species occupy wide ranges in which small, local populations become geographically isolated from the rest of the species. The allele frequencies within these small populations are more susceptible to genetic drift. Because this is a random process, small isolated populations tend to be more genetically disparate in relation to other populations. This may occur in the following two ways.

Bottleneck Effect Changes in population size may influence genetic drift via the **bottleneck effect** (Figure 25.17). In nature, a population can be reduced dramatically in size by events such as earthquakes, floods, drought, or human destruction of habitat. Such events may randomly eliminate most of the members of the population without regard to its genetic composition. The initial bottleneck may be greatly influenced by genetic drift because the surviving members may have allele frequencies that differ from those of the original population. In addition, allele frequencies are expected to drift substantially during the generations when the population size is small. In extreme cases, alleles may even be eliminated. Eventually, the population with the bottleneck may regain its original size. However, the new population has less genetic variation than the original large population. As an example, the African cheetah population lost a substantial amount of its genetic variation due to a bottleneck effect. DNA analysis by population geneticists has suggested that a severe bottleneck occurred approximately 10,000 to 12,000 years ago, when the population size was dramatically reduced. The population eventually rebounded, but the bottleneck significantly decreased the genetic variation.


Large, genetically diverse population
Bottleneck: Fewer individuals, less diversity
Large, less genetically diverse population

(a) Bottleneck effect

(b) An African cheetah

FIGURE 25.17 **The bottleneck effect, an example of genetic drift. (a)** A representation of the bottleneck effect. Note that the genetic variation denoted by the green balls has been lost. **(b)** The African cheetah. The modern species has low genetic variation due to a genetic bottleneck that is thought to have occurred about 10,000 to 12,000 years ago.

Concept check: What is happening at the bottleneck? Describe the effect of genetic drift during the bottleneck.

Founder Effect Geography and population size may also influence genetic drift via the **founder effect.** The key difference between the bottleneck effect and the founder effect is that the founder effect involves migration; a small group of individuals separates from a larger population and establishes a colony in a new location. For example, a few individuals may migrate from a large continental population and become the founders of an island population. The founder effect has two important consequences. First, the founding population is expected to have less genetic variation than the original population from which it was derived. Second, as a matter of chance, the allele frequencies in the founding population may differ markedly from those of the original population.

Population geneticists have studied many examples of isolated populations that were started from a few members of another population. In the 1960s, Victor McKusick studied allele frequencies in the Old Order Amish of Lancaster County, Pennsylvania. At that time, this was a group of about 8000 people, descended from just three couples who immigrated to the United States in the 1700s. Among this population of 8000, a genetic disease known as the Ellis-van Creveld syndrome (a recessive form of dwarfism) was found at a frequency of 0.07, or 7%. By comparison, this disorder is extremely rare in other human populations, even the population from which the founding members had originated. The high frequency of dwarfism in the Lancaster County population is a chance occurrence due to the founder effect. The recessive allele can be traced back to one couple who came to the area in 1744.

25.4 REVIEWING THE KEY CONCEPTS

- Genetic drift is a change in allele frequency due to chance fluctuations. Over the long run, it often results in allele fixation or loss. The effect of genetic drift is greater in small populations (see Figure 25.16).
- Two examples of genetic drift are the bottleneck effect and the founder effect (see Figure 25.17).

25.4 COMPREHENSION QUESTIONS

1. Genetic drift is
   a. a change in allele frequencies due to random fluctuations.
   b. likely to result in allele loss or fixation over the long run.
   c. more pronounced in smaller populations.
   d. all of the above.
2. Which of the following types of genetic drift involve the migration of a population from one location to another?
   a. The bottleneck effect
   b. The founder effect
   c. Both the a and b
   d. None of the above

25.5 MIGRATION

We have just seen how migration to a new location by a relatively small group can result in a population with an altered genetic composition. In addition, migration between two different established populations can alter allele frequencies. For example, a species of birds may occupy two geographic regions that are separated by a large body of water. On rare occasions, the prevailing winds may allow birds from the western population to fly over this body of water and become members of the eastern population. If the two populations have different allele frequencies and if migration occurs in sufficient numbers, this migration may alter the allele frequencies in the eastern population.

After migration has occurred, the new (eastern) population is called a **conglomerate.** To calculate the allele frequencies in the conglomerate, we need two kinds of information. First, we must know the original allele frequencies in the donor and recipient populations. Second, we must know the proportion of the conglomerate population that is due to migrants. With these data, we begin by calculating the change in allele frequency in the conglomerate population using the following equation:

$$\Delta p_C = m(p_D - p_R)$$

where

Δp_C is the change in allele frequency in the conglomerate population

p_D is the allele frequency in the donor population

p_R is the allele frequency in the original recipient population

m is the proportion of migrants that make up the conglomerate population

$$m = \frac{\text{number of migrants in the conglomerate population}}{\text{total number of individuals in the conglomerate population}}$$

As an example, let's suppose the allele frequency of *A* is 0.7 in the donor population and 0.3 in the recipient population. A group of 20 individuals migrates and joins the recipient population, which originally had 80 members. Thus,

$$\begin{aligned} m &= \frac{20}{20 + 80} \\ &= 0.2 \end{aligned}$$

$$\begin{aligned} \Delta p_C &= m(p_D - p_R) \\ &= 0.2(0.7 - 0.3) \\ &= 0.08 \end{aligned}$$

We can now calculate the allele frequency in the conglomerate:

$$\begin{aligned} p_C &= p_R + \Delta p_C \\ &= 0.3 + 0.08 = 0.38 \end{aligned}$$

Therefore, in the conglomerate population, the allele frequency of *A* has changed from 0.3 (its value before migration) to 0.38. This increase in allele frequency arises from the higher allele frequency of *A* in the donor population. **Gene flow** is the phenomenon in which individuals migrate from one population to another population and the migrants are able to breed successfully with the members of the recipient population. Gene flow depends not only on migration, but also on the ability of the migrants' alleles to be passed to subsequent generations.

In our previous example, we considered the consequences of a unidirectional migration from a donor to a recipient population. In nature, it is common for individuals to migrate in both directions. What are the main consequences of bidirectional migration? Depending on its rate, migration tends to reduce differences in allele frequencies between neighboring populations. Populations that frequently mix their gene pools via migration tend to have similar allele frequencies, whereas isolated populations are expected to be more disparate. By comparison, genetic drift, which we considered earlier, tends to make local populations more disparate from each other.

Migration can also enhance genetic diversity within a population. As mentioned, new mutations are relatively rare events. Therefore, a particular mutation may arise only in one population. Migration may then introduce this new allele into neighboring populations.

25.5 REVIEWING THE KEY CONCEPTS

- Migration can alter allele frequencies. It tends to reduce differences in allele frequencies between neighboring populations and increase genetic diversity within a population.

25.5 COMPREHENSION QUESTIONS

1. Gene flow depends on
   a. migration.
   b. the ability of migrant alleles to be passed to subsequent generations.
   c. genetic drift.
   d. both a and b.

25.6 NONRANDOM MATING

As mentioned earlier, one of the conditions required to establish the Hardy-Weinberg equilibrium is random mating. This means that individuals choose their mates irrespective of their genotypes and phenotypes. In many cases, particularly in human populations, this condition is violated frequently.

When mating is nonrandom in a population, the process is called **assortative mating.** Positive assortative mating occurs when individuals with similar phenotypes choose each other as mates. The opposite situation, where dissimilar phenotypes mate preferentially, is called negative assortative mating. In addition, individuals may choose a mate that is part of the same genetic lineage. The mating of two genetically related individuals, such as cousins, is called **inbreeding.** This is also termed consanguinity. Inbreeding sometimes occurs in human societies and is more likely to take place in nature when population size becomes very limited. In Chapter 26, we will examine how inbreeding is a useful strategy for developing agricultural breeds or strains with desirable characteristics. Conversely, **outbreeding,** which involves mating between unrelated individuals, can create hybrids that are heterozygous for many genes.

In the absence of other evolutionary processes, inbreeding and outbreeding do not affect allele frequencies in a population. However, these patterns of mating do disrupt the balance of genotypes that is predicted by the Hardy-Weinberg equation. Let's first consider inbreeding in a family pedigree. **Figure 25.18** illustrates a human pedigree involving a mating between cousins. Individuals III-2 and III-3 are cousins and have produced the daughter labeled IV-1. She is said to be inbred, because her parents are genetically related to each other.

Inbreeding involves a smaller gene pool because the parents are related genetically. In the 1940s, Gustave Malécot developed

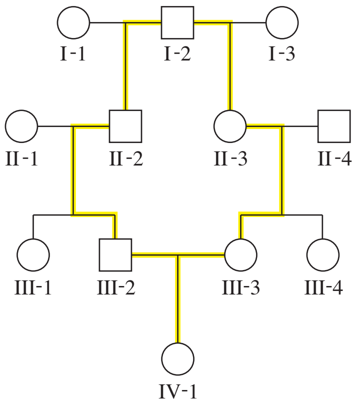


FIGURE 25.18 **A human pedigree containing inbreeding.** Individual IV-1 is the result of inbreeding because her parents are related.

Concept check: How does inbreeding affect the likelihood of expressing recessive traits? Explain.

methods to quantify the degree of inbreeding. The **inbreeding coefficient (*F*)** is the probability that two alleles in a particular individual will be identical for a given gene because both copies are due to descent from a common ancestor. An inbreeding coefficient (F) can be computed by analyzing the degree of relatedness within a pedigree.

As an example, let's determine the inbreeding coefficient for individual IV-1. To begin this problem, we must first identify all of this individual's common ancestors. A common ancestor is anyone who is an ancestor to both of an individual's parents. In Figure 25.18, IV-1 has one common ancestor, I-2, her great-grandfather. I-2 is the grandfather of III-2 and III-3.

Our next step is to determine the inbreeding paths. An inbreeding path for an individual is the shortest path through the pedigree that includes both parents and the common ancestor. In a pedigree, there is an inbreeding path for each common ancestor. The length of each inbreeding path is calculated by adding together all of the individuals in the path except the individual of interest. In this case, there is only one path because IV-1 has only one common ancestor. To add the members of the path, we begin with individual IV-1, but we do not count her. We then move to her father (III-2); to her grandfather (II-2); to I-2, her great-grandfather (the common ancestor); back down to her other grandmother (II-3); and finally to her mother (III-3). This path has five members. Finally, to calculate the inbreeding coefficient, we use the following formula:

$$F = \Sigma(1/2)^n(1 + F_A)$$

where

F is the inbreeding coefficient of the individual of interest

n is the number of individuals in the inbreeding path, excluding the inbred offspring

F_A is the inbreeding coefficient of the common ancestor

Σ indicates that we add together $(1/2)^n(1 + F_A)$ for each inbreeding path

In this case, there is only one common ancestor and, therefore, only one inbreeding path. Also, we do not know anything about the heritage of the common ancestor, so we assume that F_A is zero. Thus, in our example of Figure 25.18,

$$F = \Sigma(1/2)^n(1 + 0)$$
$$= (1/2)^5 = 1/32 = 3.125\%$$

What does this value mean? The inbreeding coefficient, 3.125%, tells us the probability that a gene in the inbred individual (IV-1) is homozygous due to its inheritance from a common ancestor (I-2). In this case, each gene in individual IV-1 has a 3.125% chance of being homozygous because the individual has inherited the same allele twice from her great-grandfather (I-2), once through each parent.

As an example, let's suppose that the common ancestor (I-2) is heterozygous for the gene involved with cystic fibrosis (CF). His genotype would be *Cc*, where *c* is the recessive allele that causes CF. Since F is 3.125%, there is a 3.125% probability that the inbred individual (IV-1) is homozygous (*CC* or *cc*) for this gene because she has inherited both copies from her great-grandfather. She has a 1.56% probability of inheriting both normal alleles (*CC*) and a 1.56% probability of inheriting both mutant alleles (*cc*). The inbreeding coefficient is denoted by the letter F (for Fixation) because it is the probability that an allele will be fixed in the homozygous condition. The term *fixation* signifies that the homozygous individual can pass only one type of allele to their offspring.

In other pedigrees, an individual may have two or more common ancestors. In this case, the inbreeding coefficient F is calculated as the sum of the inbreeding paths. Such an example is described in Solved Problem S2 at the end of the chapter.

The effects of inbreeding and outbreeding can also be considered within a population. For example, let's consider the situation in which the frequency of allele $A = p$ and the frequency of allele $a = q$. In a given population, the genotype frequencies are determined in the following way:

$p^2 + fpq$ equals the frequency of *AA*

$2pq(1 - f)$ equals the frequency of *Aa*

$q^2 + fpq$ equals the frequency of *aa*

where f is a measure of how much the genotype frequencies deviate from Hardy-Weinberg equilibrium due to nonrandom mating. The value of f ranges from -1 to $+1$. When inbreeding occurs, the value is greater than zero. When outbreeding occurs, the value is less than zero.

As an example, let's suppose that $p = 0.8$, $q = 0.2$, and $f = 0.25$. We can calculate the frequencies of the *AA*, *Aa*, and *aa* genotypes under these conditions as follows:

$$AA = p^2 + fpq = (0.8)^2 + (0.25)(0.8)(0.2) = 0.68$$
$$Aa = 2pq(1 - f) = 2(0.8)(0.2)(1 - 0.25) = 0.24$$
$$aa = q^2 + fpq = (0.2)^2 + (0.25)(0.8)(0.2) = 0.08$$

There will be 68% *AA* homozygotes, 24% heterozygotes, and 8% *aa* homozygotes. If mating had been random (i.e., $f = 0$), the genotype frequencies of *AA* would be p^2, which equals 64%, and *aa* would be q^2, which equals 4%. The frequency of heterozygotes would be $2pq$, which equals 32%. When comparing these numbers, we see that inbreeding raises the proportions of homozygotes and decreases the proportion of heterozygotes. In natural populations, the value of f tends to become larger as a population becomes smaller, because each individual has a more limited choice in mate selection.

What are the consequences of inbreeding in a population? From an agricultural viewpoint, it results in a higher proportion of homozygotes, which may exhibit a desirable trait. For example, an animal breeder may use inbreeding to produce animals that are larger because they have become homozygous for alleles promoting larger size. On the negative side, many genetic diseases are inherited in a recessive manner (see Chapter 23). For these disorders, inbreeding increases the likelihood that an individual will be homozygous and therefore afflicted with the disease. Also, in natural populations, inbreeding lowers the mean fitness of the population if homozygous offspring have a lower fitness value. This can be a serious problem as natural populations become smaller due to human habitat destruction. As the population shrinks, inbreeding becomes more likely because individuals have fewer potential mates from which to choose. The inbreeding, in turn, produces homozygotes that are less fit, thereby decreasing the reproductive success of the population. This phenomenon is called **inbreeding depression.** Conservation biologists sometimes try to circumvent this problem by introducing individuals from one population into another. For example, the endangered Florida panther (*Felis concolor coryi*) suffers from inbreeding-related defects, which include poor sperm quality and quantity and morphological abnormalities. To help alleviate these effects, panthers of the same species from Texas have been introduced into the Florida population.

25.6 REVIEWING THE KEY CONCEPTS

- Nonrandom mating may alter the genotype frequencies that would be based on a Hardy-Weinberg equilibrium. Inbreeding results in a higher proportion of homozygotes in a population (see Figure 25.18).

25.6 COMPREHENSION QUESTIONS

1. Inbreeding refers to mating between individuals that are
   a. homozygous.
   b. heterozygous.
   c. part of the same genetic lineage.
   d. both a and c.

25.7 SOURCES OF NEW GENETIC VARIATION

In the previous sections, we primarily focused on genetic variation in which a single gene exists in two or more alleles. These simplified scenarios allow us to appreciate the general principles behind evolutionary mechanisms. As researchers have analyzed genetic variation at the molecular, cellular, and population level, however, they have come to understand that new genetic variation occurs in many ways (**Table 25.2**). Among eukaryotic species, sexual reproduction is an important way that new genetic variation occurs among offspring. In Chapters 3 and 7, we considered how independent assortment and crossing over during sexual reproduction may produce new combinations of alleles

TABLE **25.2**

Sources of New Genetic Variation That Occur in Populations

Type	Description
Independent assortment	The independent segregation of different homologous chromosomes may give rise to new combinations of alleles in offspring (see Chapter 3).
Crossing over	Recombination (crossing over) between homologous chromosomes can also produce new combinations of alleles that are located on the same chromosome (see Chapter 7).
Interspecies crosses	On occasion, members of different species may breed with each other to produce hybrid offspring. This topic is discussed in Chapter 27.
Prokaryotic gene transfer	Prokaryotic species possess mechanisms of genetic transfer such as conjugation, transduction, and transformation (see Chapter 9).
New alleles	Point mutations can occur within a gene to create single-nucleotide polymorphisms (SNPs). In addition, genes can be altered by small deletions and additions. Gene mutations are also discussed in Chapter 18.
Gene duplications	Events, such as misaligned crossovers, can add additional copies of a gene into a genome and lead to the formation of gene families (see Chapter 8).
Chromosome structure and number	Chromosome structure may be changed by deletions, duplications, inversions, and translocations. Changes in chromosome number result in aneuploid, polyploid, and alloploid offspring. These mechanisms are discussed in Chapters 8 and 27.
Exon shuffling	New genes can be created when exons of preexisting genes are rearranged to make a gene that encodes a protein with a new combination of protein domains.
Horizontal gene transfer	Genes from one species can be introduced into another species and become incorporated into that species' genome.
Changes in repetitive sequences	Short repetitive sequences are common in genomes due to the occurrence of transposable elements and due to tandem arrays. The number and lengths of repetitive sequences tend to show considerable variation in natural populations.

among different genes, thereby producing new genetic variation in the resulting offspring. Similarly, in Chapter 27, we will consider how breeding between members of different species may produce hybrid offspring that harbor new combinations of genetic material. Such hybridization events have been important in the evolution of new species, particularly those in the plant kingdom. Though prokaryotic species reproduce asexually, they also possess mechanisms for gene transfer, such as conjugation, transduction, and transformation (see Chapter 9). These mechanisms are important for fostering genetic variation among bacterial and archaeal populations.

Rare mutations in DNA may also give rise to new types of variation (see Table 25.2). As discussed earlier in this chapter (see Figure 25.3) and in Chapter 18, mutations may occur within a particular gene to create new alleles of that gene. Such allelic variation is common in natural populations. Also, as described in Chapter 8, gene duplications may create a gene family; each family member acquires independent mutations and often evolves more specialized functions. An example is the globin gene family (see Figure 8.7). Changes in chromosome structure and number are also important in the evolution of new species (see Chapter 27). In this section, we examine some additional mechanisms through which an organism can acquire new genetic variation. These include exon shuffling, horizontal gene transfer, and changes in repetitive sequences. The myriad of mechanisms that foster genetic variation underscores its profound importance in the evolution of species that are both well adapted to their native environments and successful at reproduction.

Mutations Provide the Source of Genetic Variation

As discussed in Chapters 8 and 18, mutations involve changes in gene sequences, chromosome structure, and/or chromosome number. Mutations are random events that occur spontaneously at a low rate or are caused by mutagens at a higher rate. In 1926, the Russian geneticist Sergei Chetverikov was the first to suggest that mutational variability provides the raw material for evolution but does not constitute a significant evolutionary change. In other words, mutation can provide new alleles to a population but does not substantially alter allele frequencies. Chetverikov proposed that populations in nature absorb mutations like a sponge and retain them in a heterozygous condition, thereby providing a source of variability for future change.

Population geneticists often consider how new mutations affect the survival and reproductive potential of the individual that inherits them. A new mutation may be deleterious, neutral, or beneficial, depending on its effect. For genes that encode proteins, the effects of new mutations depend on their influence on protein function. Deleterious and neutral mutations are far more likely to occur than beneficial ones. For example, alleles can be altered in many different ways that render an encoded protein defective. As discussed in Chapter 18, deletions and point mutations such as frameshift mutations, missense mutations, and nonsense mutations all may cause a gene to express a protein that is nonfunctional or less functional than the wild-type protein. Also, mutations in noncoding regions can alter gene expression (refer back to Table 18.3). Neutral mutations, which are not acted upon by natural selection, occur in several different ways. For example, a neutral mutation can change the wobble base without affecting the amino acid sequence of the encoded protein, or it can be a missense mutation that has no effect on protein function. Such point mutations happen at specific sites within the coding sequence. Neutral mutations can also occur within introns, the noncoding sequences of genes. By comparison, beneficial mutations are relatively uncommon. To be advantageous, a new mutation could alter the amino acid sequence of a protein to yield a better-functioning product. Although such mutations do occur, they are expected to be very rare for a population in a stable environment.

The **mutation rate** is defined as the probability that a gene will be altered by a new mutation. The rate is typically expressed as the number of new mutations in a given gene per generation. A common value for the mutation rate is in the range of 1 in 100,000 to 1 in 1,000,000, or 10^{-5} to 10^{-6} per generation. However, mutation rates vary depending on species, cell types, chromosomal location, and gene size. Furthermore, in experimental studies, the mutation rate is usually measured by following the change of a normal (functional) gene to a deleterious (nonfunctional) allele. The mutation rate producing beneficial alleles is expected to be substantially less.

It is clear that new mutations provide genetic variability, but population geneticists also want to know how much the mutation rate affects the allele frequencies in a population. Can random mutations have a large effect on allele frequencies over time? To answer this question, let's take this simple case: a gene exists in an allele, *A*; the allele frequency of *A* is denoted by the variable p. A mutation can convert the *A* allele into a different allele called *a*. The allele frequency of *a* is designated by q. The conversion of the *A* allele into the *a* allele by mutation occurs at a rate that is designated μ. If we assume that the rate of the reverse mutation (*a* to *A*) is negligible, the increase in the frequency of the *a* allele after one generation is

$$\Delta q = \mu p$$

For example, let's consider the following conditions:

$p = 0.8$ (i.e., frequency of *A* is 80%)

$q = 0.2$ (i.e., frequency of *a* is 20%)

$\mu = 10^{-5}$ (i.e., the mutation rate of converting *A* to *a*)

$\Delta q = (10^{-5})(0.8) = (0.00001)(0.8) = 0.000008$

Therefore, in the next generation (designated $n + 1$),

$$q_{n+1} = 0.2 + 0.000008 = 0.200008$$

$$p_{n+1} = 0.8 - 0.000008 = 0.799992$$

As we can see from this calculation, new mutations do not significantly alter the allele frequencies in a single generation.

We can use the following equation to calculate the change in allele frequency after any number of generations:

$$(1 - \mu)^t = \frac{p_t}{p_0}$$

where

- μ is the mutation rate of the conversion of A to a
- t is the number of generations
- p_0 is the allele frequency of A in the starting generation
- p_t is the allele frequency of A after t generations

As an example, let's suppose that the allele frequency of A is 0.8, $\mu = 10^{-5}$, and we want to know what the allele frequency will be after 1000 generations ($t = 1000$). Plugging these values into the preceding equation and solving for p_t,

$$(1 - 0.00001)^{1000} = \frac{p_t}{0.8}$$

$$p_t = 0.792$$

Therefore, after 1000 generations, the frequency of A has dropped only from 0.8 to 0.792. Again, these results point to how slowly the occurrence of new mutations changes allele frequencies. In natural populations, the rate of new mutation is rarely a significant catalyst in shaping allele frequencies. Instead, other processes such as natural selection, genetic drift, and migration have far greater effects on allele frequencies.

New Genes Are Produced in Eukaryotes via Exon Shuffling

Sources of new genetic variation are revealed when the parts of genes that encode protein domains are compared within a single species. Many proteins, particularly those found in eukaryotic species, have a modular structure composed of two or more domains with different functions. For example, certain transcription factors have discrete domains involved with hormone binding, dimerization, and DNA binding. As described in Chapter 17, the glucocorticoid receptor has a domain that binds the hormone, a second domain that facilitates protein dimerization, and a third that allows the glucocorticoid receptor to bind to a glucocorticoid response element (GRE) next to particular genes (see Figure 17.7). By comparing the modular structure of eukaryotic proteins with the genes that encode them, geneticists have discovered that each domain tends to be encoded by one coding sequence, or exon, or by a series of two or more adjacent exons.

During the evolution of eukaryotic species, many new genes have been created by a process known as **exon shuffling,** in which an exon and its flanking introns are inserted into a gene, thereby producing a new gene that encodes a protein with an additional domain (**Figure 25.19**). This process may also involve the duplication and rearrangement of exons. Exon shuffling results in novel genes that express proteins with diverse functional modules. Such proteins can then alter traits in the organism and may be acted on by evolutionary processes such as natural selection and genetic drift.

Exon shuffling may occur by more than one mechanism. Transposable elements, which are described in Chapter 21, may promote the insertion of exons into the coding sequences of genes. Alternatively, a double crossover event could promote the insertion of an exon into another gene (this is the case in Figure 25.19). This is called nonhomologous recombination because the two regions involved in the crossover are not homologous to each other.

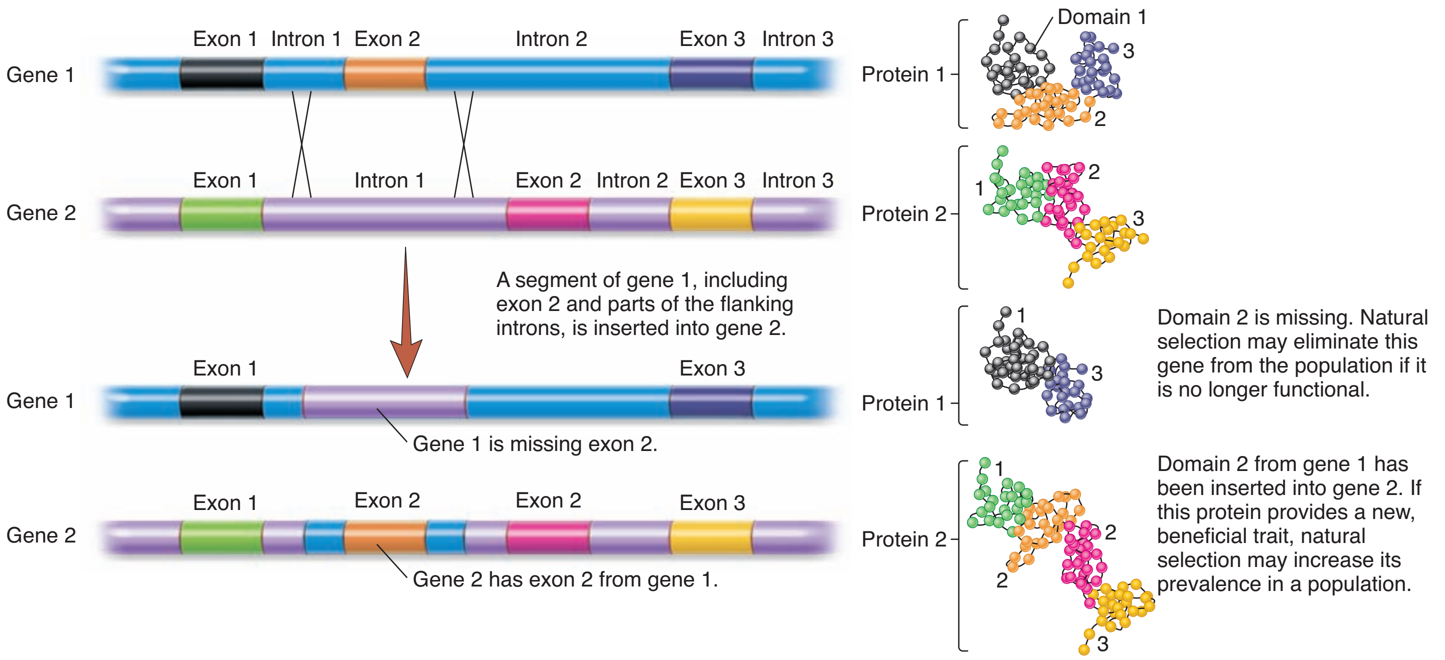
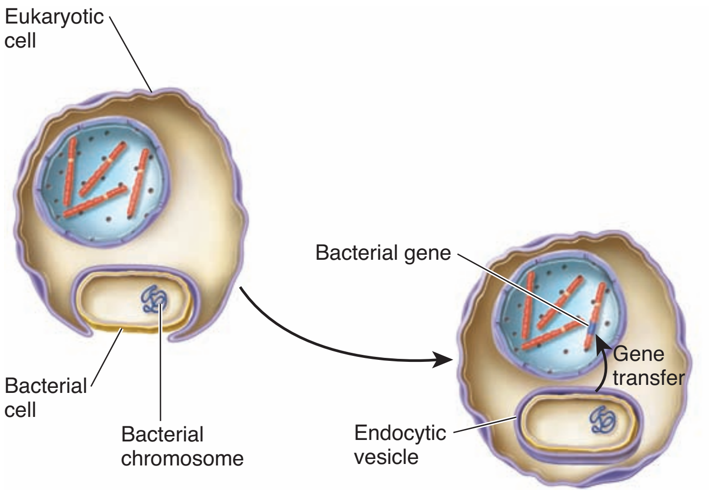

Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Intron 3
Gene 1
Protein 1
Domain 1
3
2
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Intron 3
Gene 2
Protein 2
1
2
3
A segment of gene 1, including exon 2 and parts of the flanking introns, is inserted into gene 2.
Exon 1
Exon 3
Gene 1
Protein 1
1
3
Domain 2 is missing. Natural selection may eliminate this gene from the population if it is no longer functional.
Gene 1 is missing exon 2.
Exon 1
Exon 2
Exon 2
Exon 3
Gene 2
Protein 2
1
2
2
3
Domain 2 from gene 1 has been inserted into gene 2. If this protein provides a new, beneficial trait, natural selection may increase its prevalence in a population.
Gene 2 has exon 2 from gene 1.

FIGURE 25.19 **The process of exon shuffling.** In this example, a segment of one gene containing an exon and its flanking introns has been inserted into another gene. A rare, abnormal crossing over event called nonhomologous recombination may cause this to happen. This results in proteins that have new combinations of domains and possibly new functions.

New Genes Are Acquired via Horizontal Gene Transfer

Species also accumulate genetic changes by a process called **horizontal gene transfer,** in which an organism incorporates genetic material from another organism without being the offspring of that organism. It often involves the exchange of genetic material between different species. **Figure 25.20** illustrates one possible mechanism for horizontal gene transfer. In this example, a eukaryotic cell has engulfed a bacterium by endocytosis. During the degradation of the bacterium, a bacterial gene escapes to the nucleus of the cell, where it is inserted into one of the chromosomes. In this way, a gene has been transferred from a bacterial species to a eukaryotic species. By analyzing gene sequences among many different species, researchers have discovered that horizontal gene transfer is a common phenomenon. This process can occur from prokaryotes to eukaryotes, from eukaryotes to prokaryotes, between different species of prokaryotes, and between different species of eukaryotes.

Gene transfer among bacterial species is relatively widespread. As discussed in Chapter 9, bacterial species may carry out three natural mechanisms of gene transfer known as conjugation, transduction, and transformation. By analyzing the genomes of bacterial species, scientists have determined that many genes within a given bacterial genome are derived from other species via horizontal gene transfer. Genome studies have suggested that as much as 20% to 30% of the variation in the genetic composition of modern prokaryotic species can be attributed to this process. For example, roughly 17% of the genes of *E. coli* and *Salmonella typhimurium* have been acquired from other species via horizontal gene transfer during the past 100 million years. The roles of these acquired genes are quite varied, though they commonly involve functions that are readily acted on by natural selection. These include genes that confer antibiotic resistance, the ability to degrade toxic compounds, and pathogenicity (the ability to cause disease).



ONLINE ANIMATION

FIGURE 25.20 **Horizontal gene transfer from a bacterium to a eukaryote.** In this example, a bacterium is engulfed by a eukaryotic cell, and a bacterial gene is transferred to one of the eukaryotic chromosomes.

Genetic Variation Is Produced via Changes in Repetitive Sequences

Another source of genetic variation comes from changes in **repetitive sequences**—short sequences typically a few base pairs to a few thousand base pairs long that are repeated many times within a species' genome. Repetitive sequences usually come from two types of sources. First, transposable elements (TEs) are genetic sequences that can move from place to place in a species' genome (see Chapter 21). The prevalence and movement of TEs provide a great deal of genetic variation between species and within a single species. In certain eukaryotic species, TEs have become fairly abundant (see Table 21.5).

A second type of repetitive sequence is nonmobile and involves short sequences that are tandemly repeated. In a **microsatellite** (also called short tandem repeats, STRs), the repeat unit is usually 1 to 6 bp long, and the whole tandem repeat is less than a couple hundred base pairs in length. For example, the most common microsatellite encountered in humans is a sequence $(CA)_N$, where N may range from 5 to more than 50. In other words, this dinucleotide sequence can be tandemly repeated 5 to 50 or more times. The $(CA)_N$ microsatellite is found, on average, about every 10,000 bases in the human genome. In a **minisatellite,** the repeat unit is typically 6 to 80 bp in length, and the size of the minisatellite ranges from 1 kbp to 20 kbp. An example of a minisatellite in humans is telomeric DNA. In a human sperm cell, for example, the repeat unit is 6 bp and the size of a telomere is about 15 kbp. (Note: Tandem repeat sequences are called satellites because they sediment away from the rest of the chromosomal DNA during equilibrium gradient centrifugation.)

Tandem repetitive sequences such as microsatellites and minisatellites tend to undergo mutation in which the number of tandem repeats changes. For example, a microsatellite with a 4-bp repeat unit and a length of 64 bp may undergo a mutation that adds three more repeat units and become 76 bp long. Micro- and minisatellites may change in length by different mechanisms. One common mechanism to explain this phenomenon is DNA strand slippage that occurs during DNA replication. As discussed in Chapter 18, this mechanism can explain trinucleotide repeat expansion (TNRE) (see Figure 18.12).

Because repetitive sequences tend to vary within a population, they have become a common tool that geneticists use in a variety of ways. For example, as described in Chapters 21 and 23, microsatellites are used as molecular markers to map the locations of genes (see Figure 23.7). Likewise, population geneticists analyze microsatellites or minisatellites to study variation at the population level and to determine the relationships among individuals and neighboring populations. The sizes of microsatellites and minisatellites found in closely related individuals tend to be more similar than they are in unrelated individuals. As described next, this phenomenon is the basis for DNA fingerprinting.

DNA Fingerprinting Is Used for Identification and Relationship Testing

The technique of **DNA fingerprinting,** also known as **DNA profiling,** analyzes individuals based on the occurrence of repetitive

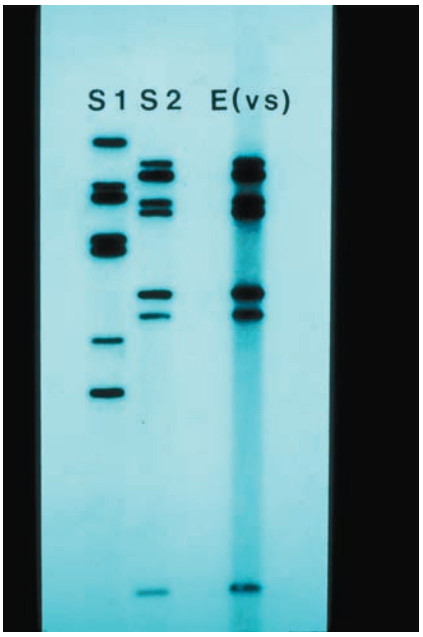


FIGURE 25.21 **A comparison of two DNA fingerprints.** The chromosomal DNA from two different individuals (Suspect 1—S1, and Suspect 2—S2) was subjected to DNA fingerprinting. The DNA evidence at a crime scene, E(vs), was also subjected to DNA fingerprinting. Following the hybridization of a radiolabeled probe, the DNA appears as a series of bands on a gel. The dissimilarity in the pattern of these bands distinguishes different individuals, much as the differences in physical fingerprint patterns can be used for identification. As seen here, S2 matches the DNA found at the crime scene.

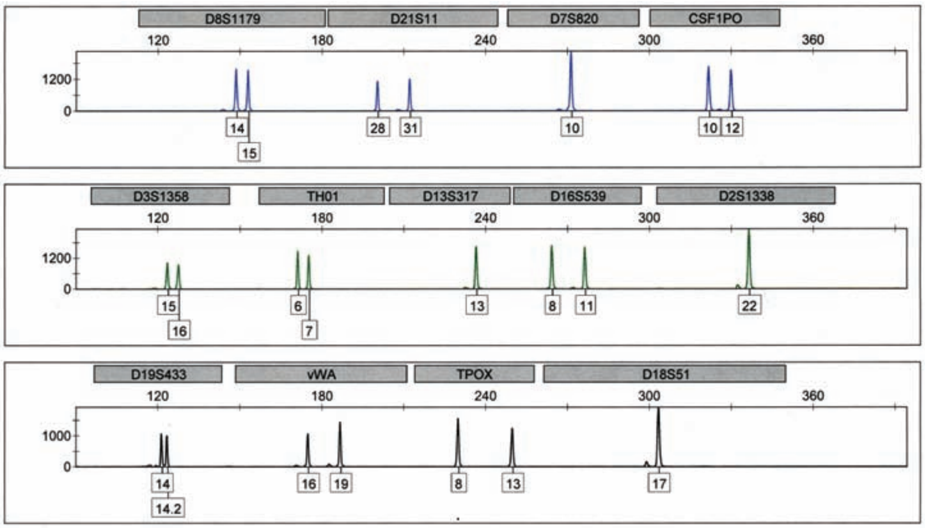
Concept check: What are two common applications of DNA fingerprinting?

sequences in their genome. When subjected to traditional DNA fingerprinting, the chromosomal DNA gives rise to a series of bands on a gel (**Figure 25.21**). The sizes and order of bands is an individual's DNA fingerprint. Like the human fingerprint, the DNA of each individual has a distinctive pattern. It is the unique pattern of these bands that makes it possible to distinguish individuals.

A comparison of the DNA fingerprints among different individuals has found two applications. First, DNA fingerprinting is used as a method of identification. In forensics, DNA fingerprinting can identify a crime suspect. In medicine, the technique can identify the type of bacterium that is causing an infection in a particular patient. A second use of DNA fingerprinting is relationship testing. Closely related individuals have more similar fingerprints than do distantly related ones (see Solved Problem S6). In humans, this is used in paternity testing. In population genetics, DNA fingerprinting can provide evidence regarding the degree of relatedness among members of a population. Such information may help geneticists determine if a population is likely to be suffering from inbreeding depression.

The development of DNA fingerprinting has relied on the identification of DNA sites that vary in length among members of a population. This naturally occurring variation causes each individual to have a unique DNA fingerprint. In the 1980s, Alec Jeffreys and his colleagues found that certain minisatellites within human chromosomes are particularly variable in their lengths. As discussed earlier, minisatellites tend to vary within populations due to changes in the number of tandem repeats at each site.

In the past decade, the technique of DNA fingerprinting has become automated, much like the automation that changed the procedure of DNA sequencing described in Chapter 19. DNA fingerprinting is now done using the technique of polymerase chain reaction (PCR), which amplifies microsatellites. Like minisatellites, microsatellites are found in multiple sites in the genome of humans and other species and vary in length among different individuals. In this procedure, the microsatellites from a sample of DNA are amplified by PCR using primers that flank the repetitive region and then separated by gel electrophoresis according to their molecular masses. As in automated DNA sequencing, the amplified microsatellite fragments are fluorescently labeled. A laser excites the fluorescent molecule within a microsatellite, and a detector records the amount of fluorescence emission for each microsatellite. As shown in **Figure 25.22**, this type of DNA fingerprint yields a series of peaks, each peak having a characteristic molecular mass. In this automated approach, the pattern of peaks rather than bands constitutes an individual's DNA fingerprint.



ONLINE ANIMATION

FIGURE 25.22 **Automated DNA fingerprinting.** In automated DNA fingerprinting, a sample of DNA is amplified, using primers that recognize the ends of microsatellites. The microsatellite fragments are fluorescently labeled and then separated by gel electrophoresis. The fluorescent molecules within each microsatellite are excited with a laser, and the amount of fluorescence is measured via a fluorescence detector. A printout from the detector is shown here. The gray boxes indicate the names of specific microsatellites. The peaks show the relative amounts of each microsatellite. The boxes beneath each peak indicate the number of tandem repeats in a given microsatellite. In this example, the individual is heterozygous for certain microsatellites (e.g., D8S1179) and homozygous for others (e.g., D7S820).

25.7 REVIEWING THE KEY CONCEPTS

- A variety of different mechanisms bring about genetic variation (see Table 25.2).
- Mutation rate is the probability that a gene will be altered by a new mutation.
- New genes in eukaryotes are produced by exon shuffling (see Figure 25.19).
- A species may acquire a new gene from another species via horizontal gene transfer (see Figure 25.20).
- A common source of genetic variation in populations involves changes in repetitive sequences, such as microsatellites.
- DNA fingerprinting is a technique that relies on variation in repetitive sequences within a population. It is used as a means of identification and in relationship testing (see Figures 25.21, 25.22).

25.7 COMPREHENSION QUESTIONS

1. The mutation rate is
   a. the likelihood that a new mutation will occur in a given gene.
   b. too low to substantially change allele frequencies in a population.
   c. lower for mutations that create beneficial alleles.
   d. all of the above.
2. The transfer of an antibiotic resistance gene from one bacterial species to a different species is an example of
   a. exon shuffling.
   b. horizontal gene transfer.
   c. genetic drift.
   d. migration.
3. DNA fingerprinting analyzes the DNA from individuals based on the occurrence of ____________ in their genomes.
   a. repetitive sequences
   b. abnormalities in chromosome structure
   c. specific genes
   d. viral insertions

KEY TERMS

Page 614. population genetics, gene pool
Page 615. population, local populations, demes, genetic polymorphism, polymorphism, polymorphic, monomorphic, single-nucleotide polymorphism (SNP)
Page 616. allele frequencies, genotype frequencies
Page 617. Hardy-Weinberg equilibrium, Hardy-Weinberg equation
Page 619. disequilibrium, microevolution
Page 620. natural selection
Page 621. Darwinian fitness, directional selection, mean fitness of the population ($\overline{W}$)
Page 623. stabilizing selection
Page 624. disruptive selection
Page 625. balancing selection, heterozygote advantage, overdominance, selection coefficient (s)
Page 626. negative frequency-dependent selection
Page 628. random genetic drift, genetic drift
Page 629. bottleneck effect
Page 630. founder effect, conglomerate
Page 631. gene flow, assortative mating, inbreeding, outbreeding
Page 632. inbreeding coefficient (F)
Page 633. inbreeding depression
Page 634. mutation rate
Page 635. exon shuffling
Page 636. horizontal gene transfer, repetitive sequences, microsatellite, minisatellite, DNA fingerprinting, DNA profiling

CHAPTER SUMMARY

- Population genetics is concerned with changes in genetic variation within a population over time.

25.1 Genes in Populations and the Hardy-Weinberg Equation

- All of the alleles of every gene in a population constitute the gene pool.
- For sexually reproducing organisms, a population is a group of individuals of the same species that occupy the same region and can interbreed with one another (see Figure 25.1).
- Genetic polymorphism refers to inherited traits or genes that exhibit variation in a population (see Figure 25.2).
- Single-nucleotide polymorphisms (SNPs) are the most common type of variation in genes (see Figure 25.3).
- Geneticists analyze genetic variation by determining allele and genotype frequencies.
- The Hardy-Weinberg equation can be used to calculate genotype frequencies based on allele frequencies (see Figures 25.4, 25.5).
- Deviation from Hardy-Weinberg equilibrium indicates that evolutionary change is occurring.
- The Hardy-Weinberg equation can be used to estimate the frequency of heterozygous carriers.

25.2 Overview of Microevolution

- Mutations are the source of new genetic variation. However, the occurrence of new mutations does not greatly change allele frequencies because they are relatively rare. Other factors, such as natural selection, genetic drift, migration, and nonrandom mating, may alter allele and/or genotype frequencies (see Table 25.1).

25.3 Natural Selection

- Natural selection is a process that changes allele frequencies from one generation to the next based on fitness, which is the relative reproductive successes of different genotypes.
- Directional selection favors the extreme phenotype (see Figures 25.6–25.8).
- Stabilizing selection favors individuals with intermediate phenotypes (see Figure 25.9).
- Disruptive selection favors multiple phenotypes. This may occur due to heterogeneous environments (see Figures 25.10, 25.11).
- Balancing selection results in a stable polymorphism. Examples include heterozygote advantage and negative-frequency dependent selection (see Figures 25.12, 25.13).
- The Grants observed natural selection in a finch population. The selection involved a change in beak size due to drought conditions (see Figures 25.14, 25.15).

25.4 Genetic Drift

- Genetic drift is a change in allele frequency due to chance fluctuations. Over the long run, it often results in allele fixation or loss. The effect of genetic drift is greater in small populations (see Figure 25.16).
- Two examples of genetic drift are the bottleneck effect and the founder effect (see Figure 25.17).

25.5 Migration

- Migration can alter allele frequencies. It tends to reduce differences in allele frequencies between neighboring populations and increase genetic diversity within a population.

25.6 Nonrandom Mating

- Nonrandom mating may alter the genotype frequencies that would be based on a Hardy-Weinberg equilibrium. Inbreeding results in a higher proportion of homozygotes in a population (see Figure 25.18).

25.7 Sources of New Genetic Variation

- A variety of different mechanisms bring about genetic variation (see Table 25.2).
- Mutation rate is the probability that a gene will be altered by a new mutation.
- New genes in eukaryotes are produced by exon shuffling (see Figure 25.19).
- A species may acquire a new gene from another species via horizontal gene transfer (see Figure 25.20).
- A common source of genetic variation in populations involves changes in repetitive sequences, such as microsatellites.
- DNA fingerprinting is a technique that relies on variation in repetitive sequences within a population. It is used as a means of identification and in relationship testing (see Figures 25.21, 25.22).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. The phenotypic frequency of people who cannot taste phenylthiocarbamide (PTC) is approximately 0.3. The inability to taste this bitter substance is due to a recessive allele. If we assume there are only two alleles in the population (namely, tasters, *T*, and nontasters, *t*) and that the population is in Hardy-Weinberg equilibrium, calculate the frequencies of these two alleles.

Answer: Let p = allele frequency of the taster allele and q = the allele frequency of the nontaster allele. The frequency of nontasters is 0.3. This is the frequency of the genotype *tt*, which in this case is equal to q^2:

$$q^2 = 0.3$$

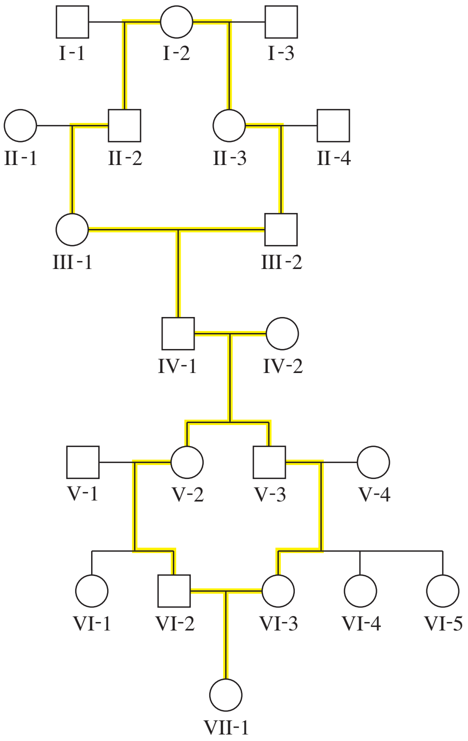
To determine the frequency q of the nontaster allele, we take the square root of both sides of this equation:

$$q = 0.55$$

With this value, we can calculate the frequency p of the taster allele:

$$p = 1 - q$$
$$= 1 - 0.55 = 0.45$$

S2. In the pedigree shown here, answer the following questions with regard to individual VII-1:



A. Who are the common ancestors of her parents?

B. What is the inbreeding coefficient?

Answer:

A. The common ancestors are IV-1 and IV-2. They are the grandparents of VI-2 and VI-3, who are the parents of VII-1.

B. The inbreeding coefficient is calculated using the formula

$$F = \Sigma(1/2)^n(1 + F_A)$$

In this case, there are two common ancestors, IV-1 and IV-2. Also, IV-1 is inbred, because I-2 is a common ancestor to both of IV-1's parents. The first step is to calculate F_A, the inbreeding coefficient for this common ancestor. The inbreeding path for IV-1 contains five people: III-1, II-2, I-2, II-3, and III-2. Therefore,

$$n = 5$$

$$F_A = (1/2)^5 = 0.03$$

Now we can calculate the inbreeding coefficient for VII-1. Each inbreeding path contains five people: VI-2, V-2, IV-1, V-3, and VI-3; and VI-2, V-2, IV-2, V-3, and VI-3. Thus,

$$F = (1/2)^5\,(1 + 0.03) + (1/2)^5\,(1 + 0)$$

$$= 0.032 + 0.031 = 0.063$$

S3. The Hardy-Weinberg equation provides a way to predict genotype frequency based on allele frequency. In the case of mammals, males are hemizygous for X-linked genes, whereas females have two copies. Among males, the frequency of any X-linked trait equals the frequency of males with the trait. For example, if an allele frequency for an X-linked disease-causing allele was 5%, then 5% of all males would be affected with the disorder. Female genotype frequencies are computed using the Hardy-Weinberg equation.

As a specific example, let's consider the human X-linked trait known as hemophilia A (see Chapter 23 for a description of this disorder). In human populations, the allele frequency of the hemophilia A allele is approximately 1 in 10,000, or 0.0001. The other allele for this gene is the normal allele. Males can be affected or unaffected, whereas females can be affected, unaffected carriers, or unaffected noncarriers.

A. What are the allele frequencies for the mutant and normal allele in the human population?

B. Among males, what is the frequency of affected individuals?

C. Among females, what is the frequency of affected individuals and heterozygous carriers?

D. Within a population of 100,000 people, what is the expected number of affected males? In this same population, what is the expected number of carrier females?

Answer: Let p represent the normal allele and q represent the allele that causes hemophilia.

A. X^H normal allele, frequency $= 0.9999 = p$

X^h hemophilia allele, frequency $= 0.0001 = q$

B. X^hY genotype frequency of affected males $= q = 0.0001$

C. X^hX^h genotype frequency of affected females $= q^2 = (0.0001)^2 = 0.00000001$

X^HX^h genotype frequency of carrier females $= 2pq = 2(0.9999)(0.0001) = 0.0002$

D. We assume this population is composed of 50% males and 50% females.

Number of affected males $= 50{,}000 \times 0.0001 = 5$

Number of carrier females $= 50{,}000 \times 0.0002 = 10$

S4. For a two-allele situation, the Hardy-Weinberg equation can also be written as

$(p + q)^2 = 1$ (Note: The number 2 in this equation reflects the idea that the genotype is due to the inheritance of two alleles, one from each parent.)

This equation can be expanded to include three or more alleles. For example, let's consider a situation in which a gene exists as three alleles: *A1*, *A2*, and *A3*. The allele frequency of *A1* is designated by the letter p, *A2* by the letter q, and *A3* by the letter r. Under these circumstances, the Hardy-Weinberg equation becomes

$$(p + q + r)^2 = 1$$

$$p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

where

p^2 is the genotype frequency of *A1A1*

q^2 is the genotype frequency of *A2A2*

r^2 is the genotype frequency of *A3A3*

$2pq$ is the genotype frequency of *A1A2*

$2pr$ is the genotype frequency of *A1A3*

$2qr$ is the genotype frequency of *A2A3*

Now here is the question. As discussed in Chapter 5, the gene that affects human blood type can exist in three alleles. In a Japanese population, the allele frequencies are

$I^A = 0.28$

$I^B = 0.17$

$i = 0.55$

Based on these allele frequencies, calculate the different possible genotype frequencies and blood type frequencies.

Answer: If we let p represent I^A, q represent I^B, and r represent i, then

p^2 is the genotype frequency of I^AI^A,
which is type A blood $= (0.28)^2 = 0.08$

q^2 is the genotype frequency of I^BI^B,
which is type B blood $= (0.17)^2 = 0.03$

r^2 is the genotype frequency of ii,
which is type O blood $= (0.55)^2 = 0.30$

$2pq$ is the genotype frequency of I^AI^B,
which is type AB blood $= 2(0.28)(0.17) = 0.09$

$2pr$ is the genotype frequency of I^Ai,
which is type A blood $= 2(0.28)(0.55) = 0.31$

$2qr$ is the genotype frequency of I^Bi,
which is type B blood $= 2(0.17)(0.55) = 0.19$

Type A $= 0.08 + 0.31 = 0.39$, or 39%

Type B $= 0.03 + 0.19 = 0.22$, or 22%

Type O $= 0.30$, or 30%

Type AB $= 0.09$, or 9%

S5. Let's suppose that pigmentation in a species of insect is controlled by a single gene existing in two alleles, *D* for dark and *d* for light. The heterozygote *Dd* is intermediate in color. In a heterogeneous environment, the allele frequencies are $D = 0.7$ and $d = 0.3$. This polymorphism is maintained because the environment contains some dimly lit forested areas and some sunny fields. During a hurricane, a group of 1000 insects is blown to a completely sunny area. In this environment, the fitness values are $DD = 0.3$, $Dd = 0.7$, and $dd = 1.0$. Calculate the allele frequencies in the next generation.

Answer: The first step is to calculate the mean fitness of the population:

$$p^2W_{DD} + 2pqW_{Dd} + q^2W_{dd} = \overline{W}$$

$$\overline{W} = (0.7)^2(0.3) + 2(0.7)(0.3)(0.7) + (0.3)^2(1.0)$$

$$= 0.15 + 0.29 + 0.09 = 0.53$$

After one generation of selection, we get

$$\text{Allele frequency of } D\text{: } p_D = \frac{p^2W_{DD}}{\overline{W}} + \frac{pqW_{Dd}}{\overline{W}}$$

$$= \frac{(0.7)^2(0.3)}{0.53} + \frac{(0.7)(0.3)(0.7)}{0.53}$$

$$= 0.55$$

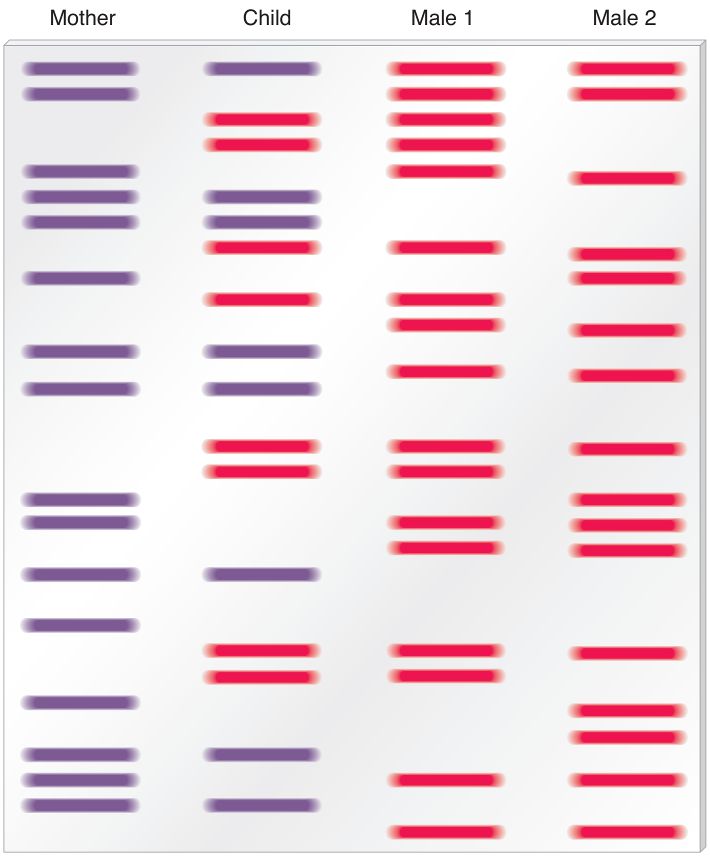
$$\text{Allele frequency of } d\text{: } q_d = \frac{q^2W_{dd}}{\overline{W}} + \frac{pqW_{Dd}}{\overline{W}}$$

$$= \frac{(0.3)^2(1.0)}{0.53} + \frac{(0.7)(0.3)(0.7)}{0.53}$$

$$= 0.45$$

After one generation, the allele frequency of *D* has decreased from 0.7 to 0.55, while the frequency of *d* has increased from 0.3 to 0.45.

S6. An important application of DNA fingerprinting is relationship testing. Persons who are related genetically have some bands or peaks in common. The number they share depends on the closeness of their genetic relationship. For example, an offspring is expected to receive half of his or her minisatellites from one parent and the rest from the other. The following diagram schematically shows a traditional DNA fingerprint of an offspring, mother, and two potential fathers.



In paternity testing, the offspring's DNA fingerprint is first compared with that of the mother. The bands that the offspring has in common with the mother are depicted in purple. The bands that are not similar between the offspring and the mother must have been inherited from the father. These bands are depicted in red. Which male could be the father?

Answer: Male 2 does not have many of the paternal bands. Therefore, he can be excluded as being the father of this child. However, male 1 has all of the paternal bands. He is very likely to be the father.

Geneticists can calculate the likelihood that the matching bands between the offspring and a prospective father could occur just as a matter of random chance. To do so, they must analyze the frequency of each band in a reference population (e.g., Caucasians living in the United States). For example, let's suppose that DNA fingerprinting analyzed 40 bands. Of these, 20 bands matched with the mother and 20 bands matched with a prospective father. If the probability of each of these bands in a reference population was 1/4, the likelihood of such a match occurring by random chance would be $(1/4)^{20}$, or roughly 1 in 1 trillion. Therefore, a match between two samples is rarely a matter of random chance.

Conceptual Questions

C1. What is the gene pool? How is a gene pool described in a quantitative way?

C2. In genetics, what does the term *population* mean? Pick any species you like and describe how its population might change over the course of many generations.

C3. What is a genetic polymorphism? What is the source of genetic variation?

C4. For each of the following, state whether it is an example of an allele, genotype, and/or phenotype frequency:

A. Approximately 1 in 2500 Caucasians is born with cystic fibrosis.

B. The percentage of carriers of the sickle cell allele in West Africa is approximately 13%.

C. The number of new mutations for achondroplasia, a genetic disorder, is approximately 5×10^{-5}.

C5. The term *polymorphism* can refer to both genes and traits. Explain the meaning of a polymorphic gene and a polymorphic trait. If a gene is polymorphic, does the trait that the gene affects also have to be polymorphic? Explain why or why not.

C6. Cystic fibrosis (CF) is a recessive autosomal trait. In certain Caucasian populations, the number of people born with this disorder is about 1 in 2500. Assuming Hardy-Weinberg equilibrium for this trait:

A. What are the frequencies for the normal and CF alleles?

B. What are the genotype frequencies of homozygous normal, heterozygous, and homozygous affected individuals?

C. Assuming random mating, what is the probability that two phenotypically unaffected heterozygous carriers will choose each other as mates?

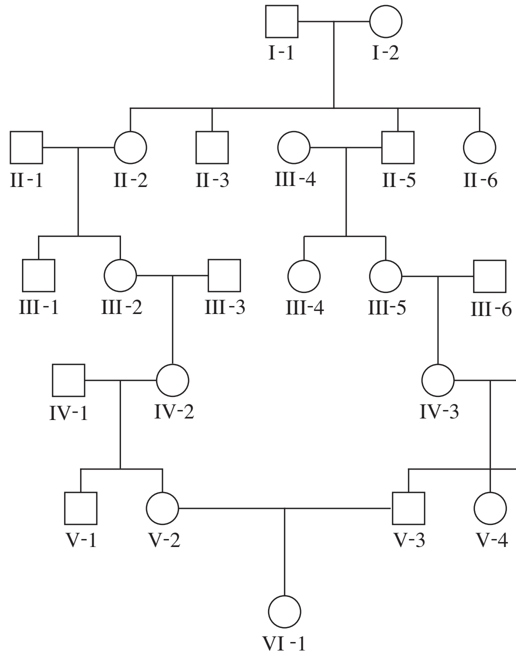
C7. Does inbreeding affect allele frequencies? Why or why not? How does it affect genotype frequencies? With regard to rare recessive diseases, what are the consequences of inbreeding in human populations?

C8. For a gene existing in two alleles, what are the allele frequencies when the heterozygote frequency is at its maximum value, assuming a Hardy-Weinberg equilibrium? What if there are three alleles?

C9. In a population, the frequencies of two alleles are $B = 0.67$ and $b = 0.33$. The genotype frequencies are $BB = 0.50$, $Bb = 0.37$, and $bb = 0.13$. Do these numbers suggest inbreeding? Explain why or why not.

C10. The ability to roll your tongue is inherited as a recessive trait. The frequency of the rolling allele is approximately 0.6, and the dominant (nonrolling) allele is 0.4. What is the frequency of individuals who can roll their tongues?

C11. Using the pedigree shown here, answer the following questions for individual VI-1:



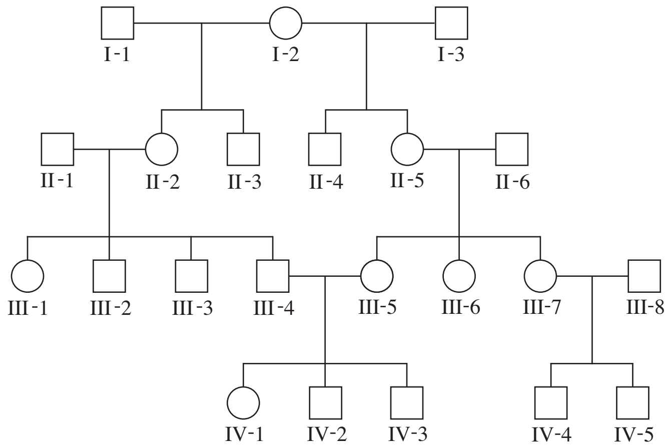
A. Is this individual inbred?

B. If so, who are her common ancestor(s)?

C. Calculate the inbreeding coefficient for VI-1.

D. Are the parents of VI-1 inbred?

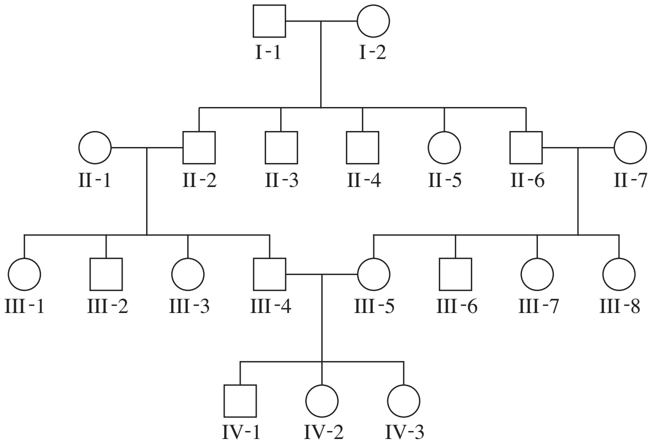
C12. A family pedigree is shown here.



A. What is the inbreeding coefficient for individual IV-3?

B. Based on the data shown in this pedigree, is individual IV-4 inbred?

C13. A family pedigree is shown here.



A. What is the inbreeding coefficient for individual IV-2? Who is/are her common ancestors?

B. Based on the data shown in this pedigree, is individual III-4 inbred?

C14. What evolutionary factors can cause allele frequencies to change and possibly lead to a genetic polymorphism? Discuss the relative importance of each type of process.

C15. In the term *genetic drift*, what is drifting? Why is this an appropriate term to describe this phenomenon?

C16. What is the difference between a random and an adaptive evolutionary process? Describe two or more examples of each. At the molecular level, explain how mutations can be random or adaptive.

C17. Let's suppose the mutation rate for converting a B allele into a b allele is 10^{-4}. The current allele frequencies are $B = 0.6$ and $b = 0.4$. How long will it take for the allele frequencies to equal each other, assuming that no genetic drift is taking place?

C18. Why is genetic drift more significant in small populations? Why does it take longer for genetic drift to cause allele fixation in large populations than in small ones?

C19. A group of four birds flies to a new location and initiates the formation of a new colony. Three of the birds are homozygous *AA*, and one bird is heterozygous *Aa*.

A. What is the probability that the *a* allele will become fixed in the population?

B. If fixation of the *a* allele occurs, how long will it take?

C. How will the growth of the population, from generation to generation, affect the answers to parts A and B? Explain.

C20. Describe what happens to allele frequencies as a result of the bottleneck effect. Discuss the relevance of this effect with regard to species that are approaching extinction.

C21. When two populations frequently intermix due to migration, what are the long-term consequences with regard to allele frequencies and genetic variation?

C22. Discuss the similarities and differences among directional, disruptive, stabilizing, and balancing selection.

C23. What is Darwinian fitness? What types of characteristics can promote high fitness values? Give several examples.

C24. What is the intuitive meaning of the mean fitness of a population? How does its value change in response to natural selection?

C25. Antibiotics are commonly used to combat bacterial and fungal infections. During the past several decades, however, antibiotic-resistant strains of microorganisms have become alarmingly prevalent. This has undermined the effectiveness of antibiotics in treating many types of infectious disease. Discuss how the following processes that alter allele frequencies may have contributed to the emergence of antibiotic-resistant strains:

A. Random mutation

B. Genetic drift

C. Natural selection

C26. With regard to genetic drift, are the following statements true or false? If a statement is false, explain why.

A. Over the long run, genetic drift leads to allele fixation or loss.

B. When a new mutation occurs within a population, genetic drift is more likely to cause the loss of the new allele rather than the fixation of the new allele.

C. Genetic drift promotes genetic diversity in large populations.

D. Genetic drift is more significant in small populations.

C27. Two populations of antelope are separated by a mountain range. The antelope are known to occasionally migrate from one population to the other. Migration can occur in either direction. Explain how migration affects the following phenomena:

A. Genetic diversity in the two populations

B. Allele frequencies in the two populations

C. Genetic drift in the two populations

C28. Do the following examples describe directional, disruptive, balancing, or stabilizing selection?

A. Polymorphisms in snail color and banding pattern as described in Figure 25.11.

B. Thick fur among mammals exposed to cold climates

C. Birth weight in humans

D. Sturdy stems and leaves among plants exposed to windy climates

Application and Experimental Questions

E1. You will need to be familiar with the techniques described in Chapter 19 to answer this question. Gene polymorphisms can be detected using a variety of cellular and molecular techniques. Which techniques would you use to detect gene polymorphisms at the following levels?

A. DNA level

B. RNA level

C. Polypeptide level

E2. You will need to understand Solved Problem S4 to answer this question. A gene for coat color in rabbits can exist in four alleles termed *C* (full coat color), c^{ch} (chinchilla), c^h (Himalayan), and *c* (albino). In a population of rabbits in Hardy-Weinberg equilibrium, the allele frequencies are

$C = 0.34$

$c^{ch} = 0.17$

$c^h = 0.44$

$c = 0.05$

Assume that *C* is dominant to the other three alleles. c^{ch} is dominant to c^h and *c*, and c^h is dominant to *c*.

A. What is the frequency of albino rabbits?

B. Among 1000 rabbits, how many would you expect to have a Himalayan coat color?

C. Among 1000 rabbits, how many would be heterozygotes with a chinchilla coat color?

E3. In a large herd of 5468 sheep, 76 animals have yellow fat, compared with the rest of the members of the herd, which have white fat. Yellow fat is inherited as a recessive trait. This herd is assumed to be in Hardy-Weinberg equilibrium.

A. What are the frequencies of the white and yellow fat alleles in this population?

B. Approximately how many sheep with white fat are heterozygous carriers of the yellow allele?

E4. The human MN blood group is determined by two codominant alleles, *M* and *N*. The following data were obtained from various human populations:

		Percentages*		
Population	**Place**	***MM***	***MN***	***NN***
Inuit	East Greenland	83.5	15.6	0.9
Navajo Indians	New Mexico	84.5	14.4	1.1
Finns	Karajala	45.7	43.1	11.2
Russians	Moscow	39.9	44.0	16.1
Aborigines	Queensland	2.4	30.4	67.2

*Data from E. B. Speiss (1990). *Genes in Populations*, 2d ed. Wiley-Liss, New York.

A. Calculate the allele frequencies in these five populations.

B. Which populations appear to be in Hardy-Weinberg equilibrium?

C. Which populations do you think have had significant intermixing due to migration?

E5. You will need to understand Solved Problem S4 before answering this question. In an island population, the following data were obtained for the numbers of people with each of the four blood types:

Type O	721
Type A	932
Type B	235
Type AB	112

Is this population in Hardy-Weinberg equilibrium? Explain your answer.

E6. In a donor population, the allele frequencies for the common (Hb^A) and sickle cell alleles (Hb^S) are 0.9 and 0.1, respectively. A group of 550 individuals migrates to a new population containing 10,000 individuals; in the recipient population, the allele frequencies are $Hb^A = 0.99$ and $Hb^S = 0.01$.

A. Calculate the allele frequencies in the conglomerate population.

B. Assuming the donor and recipient populations are each in Hardy-Weinberg equilibrium, calculate the genotype frequencies in the conglomerate population prior to further mating between the donor and recipient populations.

C. What will be the genotype frequencies of the conglomerate population in the next generation, assuming it achieves Hardy-Weinberg equilibrium in one generation?

E7. A recessive lethal allele has achieved a frequency of 0.22 due to genetic drift in a very small population. Based on natural selection, how would you expect the allele frequencies to change in the next three generations? Note: Your calculation can assume that genetic drift is not altering allele frequencies in either direction.

E8. Among a large population of 2 million gray mosquitoes, one mosquito is heterozygous for a body color gene; this mosquito has one gray allele and one blue allele. There is no selective advantage or disadvantage between gray and blue body color. All of the other mosquitoes carry the gray allele.

A. What is the probability of fixation of the blue allele?

B. If fixation happens to occur, how many generations is it likely to take?

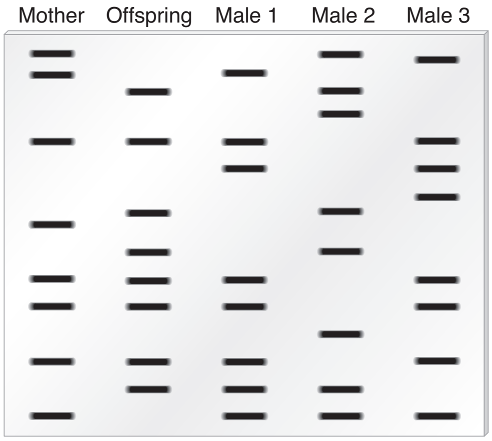
C. Qualitatively, how would the answers to parts A and B be affected if the blue allele conferred a slight survival advantage?

E9. Resistance to the poison warfarin is a genetically determined trait in rats. Homozygotes carrying the resistance allele (*WW*) have a lower fitness because they suffer from vitamin K deficiency, but heterozygotes (*Ww*) do not. However, the heterozygotes are still resistant to warfarin. In an area where warfarin is applied, the heterozygote has a survival advantage. Due to warfarin resistance, the heterozygote is also more fit than the normal homozygote (*ww*). If the relative fitness values for *Ww*, *WW*, and *ww* individuals are 1.0, 0.37, and 0.19, respectively, in areas where warfarin is applied, calculate the allele frequencies at equilibrium. How would this equilibrium be affected if the rats were no longer exposed to warfarin?

E10. Describe, in as much experimental detail as possible, how you would test the hypothesis that snail color distribution is due to predation.

E11. In the Grants' study of the medium ground finch, do you think the pattern of natural selection was directional, stabilizing, disruptive, or balancing? Explain your answer. If the environment remained dry indefinitely (for many years), what do you think would be the long-term outcome?

E12. Here are traditional DNA fingerprints of five people: a child, mother, and three potential fathers:



Which males can be ruled out as being the father? Explain your answer. If one of the males could be the father, explain the general strategy for calculating the likelihood that he could match the offspring's DNA fingerprint by chance alone. (See Solved Problem S6 before answering this question.)

E13. What is DNA fingerprinting? How can it be used in human identification?

E14. When analyzing the automated DNA fingerprints of a father and his biological daughter, a technician examined 50 peaks and found that 30 of them were a perfect match. In other words, 30 out of 50 peaks, or 60%, were a perfect match. Is this percentage too high, or would you expect a value of only 50%? Explain why or why not.

E15. What would you expect to be the minimum percentage of matching peaks in an automated DNA fingerprint for the following pairs of individuals?

A. Mother and son

B. Sister and brother

C. Uncle and niece

D. Grandfather and grandson

Questions for Student Discussion/Collaboration

1. Discuss examples of positive and negative assortative mating in natural populations, human populations, and agriculturally important species.
2. Discuss the role of mutation in the origin of genetic polymorphisms. Suppose that a genetic polymorphism involves two alleles at frequencies of 0.45 and 0.55. Describe three different scenarios to explain these observed allele frequencies. You can propose that the alleles are neutral, beneficial, or deleterious.
3. Most new mutations are detrimental, yet rare beneficial mutations can be adaptive. With regard to the fate of new mutations, discuss whether you think it is more important for natural selection to select against detrimental alleles or to select in favor of beneficial ones. Which do you think is more significant in human populations?

Answers to Comprehension Questions

25.1: c, a, a, d

25.2: a

25.3: b, a, c, d

25.4: d, b

25.5: d

25.6: c

25.7: d, b, a

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

26

CHAPTER OUTLINE

26.1 Overview of Quantitative Traits

26.2 Statistical Methods for Evaluating Quantitative Traits

26.3 Polygenic Inheritance

26.4 Heritability

26.5 Selective Breeding

Domesticated wheat. *The color of wheat ranges from a dark red to white, which is an example of a complex or quantitative trait.*

QUANTITATIVE GENETICS

In this chapter, we will examine **complex traits**—characteristics that are determined by several genes and are significantly influenced by environmental factors. Many complex traits are viewed as **quantitative traits** because they can be described numerically. In humans, quantitative traits include height, the shape of our noses, and the rate at which we metabolize food, to name a few examples. The field of genetics that studies the mode of inheritance of complex or quantitative traits is called **quantitative genetics.** Quantitative genetics is an important branch of genetics for several reasons. In agriculture, most of the key characteristics of interest to plant and animal breeders are quantitative traits. These include traits such as weight, fruit size, resistance to disease, and the ability to withstand harsh environmental conditions. As we will see later in this chapter, genetic techniques have improved our ability to develop strains of agriculturally important species with desirable quantitative traits. In addition, many human diseases are viewed as complex traits that are influenced by several genes.

Quantitative genetics is also important in the study of evolution. Many of the traits that allow a species to adapt to its environment are quantitative. Examples include the swift speed of the cheetah and the sturdiness of tree branches in windy climates. The importance of quantitative traits in the evolution of species will be discussed in Chapter 27. In this chapter, we examine how genes and the environment contribute to the phenotypic expression of quantitative traits. We will begin with an examination of quantitative traits and how to analyze them using statistical techniques. We then look at the inheritance of polygenic traits and at quantitative trait loci (QTLs)—locations on chromosomes containing genes that affect the outcome of quantitative traits. Advances in genetic mapping strategies have enabled researchers to identify these genes. Last, we look at heritability and consider various ways of calculating and modifying the genetic variation that affects phenotype.

26.1 OVERVIEW OF QUANTITATIVE TRAITS

When we compare characteristics among members of the same species, the differences are often quantitative rather than qualitative. Humans, for example, have the same basic anatomical features (two eyes, two ears, and so on), but they differ in quantitative ways. People vary with regard to height, weight, the shape of facial features, pigmentation, and many other characteristics. As shown in **Table 26.1**, quantitative traits can be categorized as anatomical, physiological, and behavioral. In addition, many human diseases exhibit characteristics and inheritance patterns

TABLE **26.1**

Types of Quantitative Traits

Trait	Examples
Anatomical traits	Height, weight, number of bristles in *Drosophila*, ear length in corn, and the degree of pigmentation in flowers and skin
Physiological traits	Metabolic traits, speed of running and flight, ability to withstand harsh temperatures, and milk production in mammals
Behavioral traits	Mating calls, courtship rituals, ability to learn a maze, and the ability to grow or move toward light
Diseases	Heart disease, hypertension, cancer, diabetes, and arthritis

analogous to those of quantitative traits. Three of the leading causes of death worldwide—heart disease, cancer, and diabetes—are considered quantitative traits.

In many cases, quantitative traits are easily measured and described numerically. Height and weight can be measured in centimeters (or inches) and kilograms (or pounds), respectively. The number of bristles on a fruit fly's body can be counted, and metabolic rate can be assessed as the amount of glucose burned per minute. Behavioral traits can also be quantified. A mating call can be evaluated with regard to its duration, sound level, and pattern. The ability to learn a maze can be described as the time and/or repetitions it takes to learn the skill. Quantitative traits, such as height and weight, are viewed as **continuous traits**—traits that do not fall into discrete categories. Some, such as bristle number in *Drosophila,* are **meristic traits**—traits that can be counted and expressed in whole numbers. Finally, certain diseases, such as diabetes, are viewed as quantitative traits even though the disease itself can be considered qualitative—either you have it or you don't. The reason why such diseases are considered quantitative traits is because the alleles of several different genes contribute to the likelihood that an individual will develop the disease. A certain threshold must be reached in which the number of disease-causing alleles results in the development of the disease. These are referred to as **threshold traits**—traits that are inherited quantitatively due to the contribution of many genes, but the traits themselves are expressed qualitatively.

Many Quantitative Traits Exhibit a Continuum of Phenotypic Variation That Follows a Normal Distribution

In Part II of this textbook, we discussed many traits that fall into discrete categories. For example, fruit flies might have white or red eyes, and pea plants might have wrinkled or smooth seeds. The alleles that govern these traits affect the phenotype in a qualitative way. In analyzing crosses involving these types of traits, each offspring can be put into a particular phenotypic category. Such attributes are called **discontinuous traits.**

In contrast, quantitative traits show a continuum of phenotypic variation within a group of individuals. For such traits, it is often impossible to place organisms into a discrete phenotypic class. For example, **Figure 26.1a** is a classic photograph from 1914 showing the range of heights of 175 students at the Connecticut Agricultural College. Though height is found at minimum and maximum values, the range of heights between these values is fairly continuous.

How do geneticists describe traits that show a continuum of phenotypes? Because most quantitative traits do not naturally fall into a small number of discrete categories, an alternative way to describe them is a **frequency distribution.** To construct a frequency distribution, the trait is divided arbitrarily into a number of convenient, discrete phenotypic categories. For example, in Figure 26.1, the range of heights is partitioned into 1-inch intervals. Then a graph is made that shows the numbers of individuals found in each of the categories.

Figure 26.1b shows a frequency distribution for the heights of students pictured in Figure 26.1a. The measurement of height is plotted along the *x*-axis, and the number of individuals who exhibit that phenotype is plotted on the *y*-axis. The values along the *x*-axis are divided into the discrete 1-inch intervals that define the phenotypic categories, even though height is essentially continuous within a group of individuals. For example, in Figure 26.1a, 22 students were between 64.5 and 65.5 inches in height, which is plotted as the point (65 inches, 22 students) on the graph in Figure 26.1b. This type of analysis can be conducted on any group of individuals who vary with regard to a quantitative trait.

The line in the frequency distribution depicts a **normal distribution,** a distribution for a large sample in which the trait of interest varies in a symmetrical way around an average value. The distribution of measurements of many biological characteristics is approximated by a symmetrical bell curve like that in Figure 26.1b. Normal distributions are common when the phenotype is determined by the cumulative effect of many small independent factors.

26.1 REVIEWING THE KEY CONCEPTS

- Quantitative traits can be categorized as anatomical, physiological, or behavioral (see Table 26.1).
- Quantitative traits often exhibit a continuum and may follow a normal distribution (see Figure 26.1).

26.1 COMPREHENSION QUESTIONS

1. Which of the following is an example of a quantitative trait?
   a. Height
   b. Rate of glucose metabolism
   c. Ability to learn a maze
   d. All of the above
2. Saying that a quantitative trait follows a continuum means that
   a. the numerical value for the trait increases with the age of the individual.
   b. environmental effects are additive.
   c. phenotypes are continuous and do not fall into discrete categories.
   d. the trait continuously changes during the life of an individual.

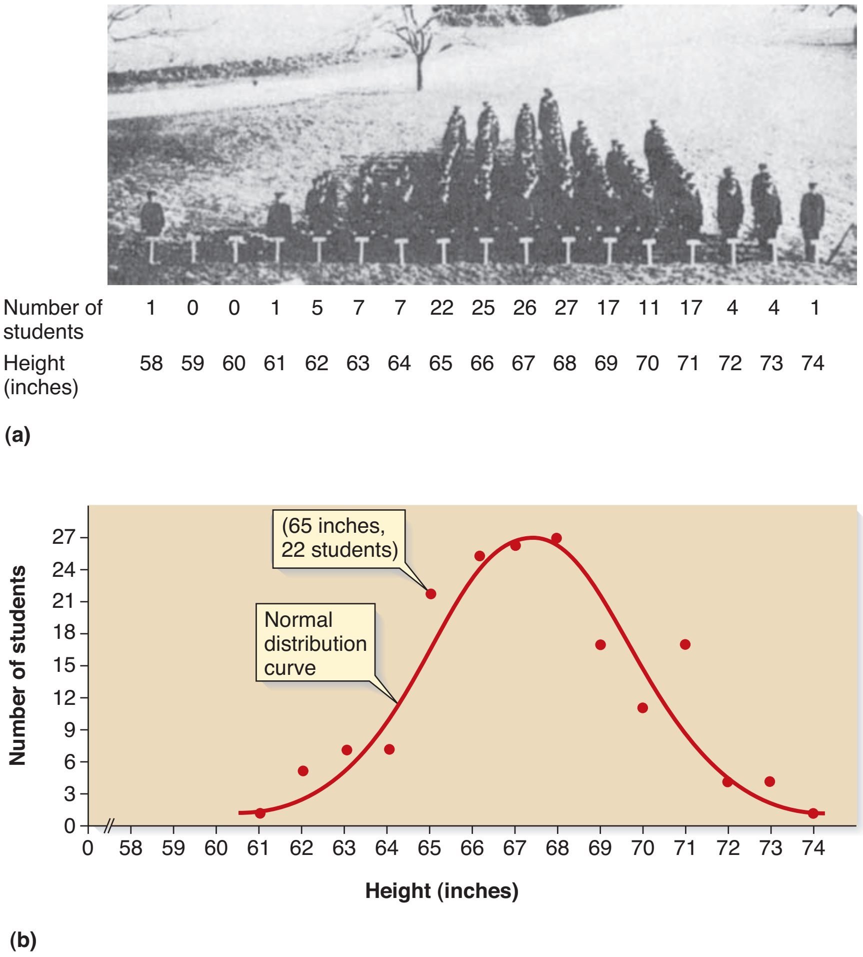


FIGURE 26.1 **Normal distribution of a quantitative trait.** **(a)** The distribution of heights in 175 students at the Connecticut Agricultural College in 1914. **(b)** A frequency distribution for the heights of students shown in (a).

Concept check: *Is height a discontinuous (discrete) trait or does it follow a continuum?*

26.2 STATISTICAL METHODS FOR EVALUATING QUANTITATIVE TRAITS

In the early 1900s, Francis Galton in England and his student Karl Pearson showed that many traits in humans and domesticated animals are quantitative in nature. To understand the underlying genetic basis of these traits, they founded what became known as the **biometric field** of genetics, which involved the statistical study of biological traits. During this period, Galton and Pearson developed various statistical tools for studying the variation of quantitative traits within groups of individuals. Many of these tools are still in use today. In this section, we will examine how statistical tools are used to analyze the variation of quantitative traits within groups.

Statistical Methods Are Used to Evaluate a Frequency Distribution Quantitatively

Statistical tools can be used to analyze a normal distribution in a number of ways. One measure you are probably familiar with is a parameter called the **mean,** which is the sum of all the values in the group divided by the number of individuals in the group. The mean is computed using the following formula:

$$\overline{X} = \frac{\Sigma X}{N}$$

where

$\overline{X}$	is the mean
ΣX	is the sum of all the values in the group
N	is the number of individuals in the group

A more generalized form of this equation can be used:

$$\overline{X} = \frac{\Sigma f_i X_i}{N}$$

where

$\overline{X}$	is the mean
$\Sigma f_i X_i$	is the sum of all the values in the group; each value in the group is multiplied by its frequency (f_i) in the group
N	is the number of individuals in the group

For example, suppose a bushel of corn had ears with the following lengths (rounded to the nearest centimeter): 15, 14, 13, 14, 15, 16, 16, 17, 15, and 15. Then

$$\overline{X} = \frac{4(15) + 2(14) + 13 + 2(16) + 17}{10}$$

$$\overline{X} = 15 \text{ cm}$$

In genetics, we are often interested in the amount of phenotypic variation that exists in a group. As we will see later in this chapter and in Chapter 27, variation lies at the heart of breeding experiments and of evolution. Without variation, selective breeding is not possible, and natural selection cannot favor one phenotype over another. A common way to evaluate variation within a population is with a statistic called the **variance,** which is a measure of the variation around the mean. The variance is the sum of the squared deviations from the mean divided by the degrees of freedom (df equals $N - 1$; see Chapter 3 for a review of degrees of freedom).

$$V_x = \frac{\Sigma f_i \, (X_i - \overline{X})^2}{N - 1}$$

where

V_X is the variance

$X_i - \overline{X}$ is the difference between each value and the mean

N is the number of observations

For example, if we use the values given previously for the lengths of ears of corn, the variance in this group is calculated as follows:

$$\Sigma f_i \, (X_i - \overline{X})^2 = 4(15 - 15)^2 + 2(14 - 15)^2 + (13 - 15)^2 + 2(16 - 15)^2 + (17 - 15)^2$$

$$\Sigma f_i \, (X_i - \overline{X})^2 = 0 + 2 + 4 + 2 + 4$$

$$\Sigma f_i \, (X_i - \overline{X})^2 = 12 \text{ cm}^2$$

$$V_X = \frac{\Sigma f_i \, (X_i - \overline{X})^2}{N - 1}$$

$$V_X = \frac{12 \text{ cm}^2}{9}$$

$$V_X = 1.33 \text{ cm}^2$$

Although variance is a measure of the variation around the mean, it is a statistic that may be difficult to understand intuitively because the variance is computed from squared deviations. For example, weight can be measured in grams; the corresponding variance is measured in square grams. Even so, variances are centrally important in the analysis of quantitative traits because they are additive under certain conditions. This means that the variances for different factors that contribute to a quantitative trait, such as genetic and environmental factors, can be added together to predict the total variance for that trait. Later, we will examine how this property is useful in predicting the outcome of genetic crosses.

To gain a more intuitive grasp of variation, we can take the square root of the variance. This statistic is called the **standard deviation** (***SD***)**.** Again, using the same values for length, the standard deviation is

$$SD = \sqrt{V_X} = \sqrt{1.33}$$

$$SD = 1.15 \text{ cm}$$

If the values in a population follow a normal distribution, it is easier to appreciate the amount of variation by considering the standard deviation. Figure 26.2 illustrates the relationship between the standard deviation and the percentages of individuals that deviate from the mean. Approximately 68% of all individuals have values within one standard deviation from the mean, either in the positive or negative direction. About 95% are within two standard deviations, and 99.7% are within three standard deviations. When a quantitative characteristic follows a normal distribution, less than 0.3% of the individuals have values that are more or less than three standard deviations from the mean of the population. In our corn example, three standard deviations equal 3.45 cm. Therefore, we expect that approximately 0.3% of the ears of corn would be less than 11.55 cm or greater than 18.45 cm, assuming that length follows a normal distribution.

Some Statistical Methods Compare Two Variables with Each Other

In many biological problems, it is useful to compare two different variables. For example, we may wish to compare the occurrence of two different phenotypic traits. Do obese animals have larger hearts? Are brown eyes more likely to occur in people with dark skin pigmentation? A second type of comparison is between traits and environmental factors. Does insecticide resistance occur more frequently in areas that have been exposed to insecticides? Is heavy body weight more prevalent in colder climates?

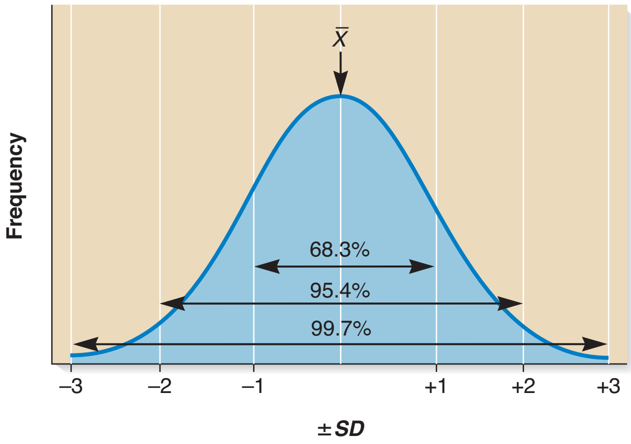


FIGURE 26.2 The relationship between the standard deviation and the proportions of individuals in a normal distribution. For example, approximately 68% of the individuals in a population are between the mean and one standard deviation (1 *SD*) above or below the mean.

Concept check: What percentage of individuals fall above two standard deviations?

Finally, a third type of comparison is between traits and genetic relationships. Do tall parents tend to produce tall offspring? Do women with diabetes tend to have brothers with diabetes?

To gain insight into such questions, a statistic known as the correlation coefficient is often applied. To calculate this statistic, we first need to determine the **covariance,** which describes the relationship between two variables within a group. The covariance is similar to the variance, except that we multiply together the deviations of two different variables rather than squaring the deviations from a single factor.

$$CoV_{(X,Y)} = \frac{\Sigma f_i[(X_i - \overline{X})\,(Y_i - \overline{Y})]}{N - 1}$$

where

$CoV_{(X,Y)}$ is the covariance between X and Y values

X_i represents the values for one variable, and $\overline{X}$ is the mean value in the group

Y_i represents the values for another variable, and $\overline{Y}$ is the mean value in that group

N is the total number of pairs of observations

As an example, let's compare the weight of cows and that of their adult female offspring. A farmer might be interested in this relationship to determine if genetic variation plays a role in the weight of cattle. The data here describe the weights at 5 years of age for 10 different cows and their female offspring.

Mother's Weight (kg)	Offspring's Weight (kg)	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X})\,(Y_i - \overline{Y})$
570	568	−26	−30	780
572	560	−24	−38	912
599	642	3	44	132
602	580	6	−18	−108
631	586	35	−12	−420
603	642	7	44	308
599	632	3	34	102
625	580	29	−18	−522
584	605	−12	7	−84
575	585	−21	−13	273
$\overline{X} = 596$	$\overline{Y} = 598$			$\Sigma = 1373$

$SD_X = 21.1$ $\quad SD_Y = 30.5$

$$CoV_{(X,Y)} = \frac{\Sigma f_i[(X_i - \overline{X})\,(Y_i - \overline{Y})]}{N - 1}$$

$$CoV_{(X,Y)} = \frac{1373}{10 - 1}$$

$$CoV_{(X,Y)} = 152.6$$

After we calculate the covariance, we can evaluate the strength of the association between the two variables by calculating a **correlation coefficient (*r*).** This value, which ranges between −1 and +1, indicates how two factors vary in relation to each other. The correlation coefficient is calculated as

$$r_{(X,Y)} = \frac{CoV_{(X,Y)}}{SD_X SD_Y}$$

A positive r value means that two factors tend to vary in the same way relative to each other; as one factor increases, the other increases with it. A value of zero indicates that the two factors do not vary in a consistent way relative to each other; the values of the two factors are not related. Finally, a negative correlation, in which the correlation coefficient is negative, indicates that the two factors tend to vary in opposite ways to each other; as one factor increases, the other decreases.

Let's use the data of 5-year weights for mother and offspring to calculate a correlation coefficient.

$$r_{(X,Y)} = \frac{152.6}{(21.1)(30.5)}$$

$$r_{(X,Y)} = 0.237$$

The result is a positive correlation between the 5-year weights of mother and offspring. In other words, the positive correlation value suggests that heavy mothers tend to have heavy offspring and lighter mothers have lighter offspring.

How do we evaluate the value of r? After a correlation coefficient has been calculated, one must consider whether the r value represents a true association between the two variables or if it could simply be due to chance. To accomplish this, we can test the hypothesis that there is no real correlation (i.e., the null hypothesis, $r = 0$). The null hypothesis is that the observed r value differs from zero due only to random sampling error. We followed a similar approach in the chi square analysis described in Chapter 3. Like the chi square value, the significance of the correlation coefficient is directly related to sample size and the degrees of freedom (*df*). In testing the significance of correlation coefficients, *df* equals $N - 2$, because two variables are involved. N equals the number of paired observations. We will reject the null hypothesis if the chi square value results in a probability that is less than 0.05 (less than 5%) or if the probability is less than 0.01 (less than 1%). These are called the 5% and 1% significance levels, respectively. **Table 26.2** shows the relationship between the r values and degrees of freedom at the 5% and 1% significance levels.

The use of Table 26.2 is valid only if several assumptions are met. First, the values of X and Y in the study must have been obtained by an unbiased sampling of the entire population. In addition, this approach assumes that the values of X and Y follow a normal distribution, like that of Figure 26.1, and that the relationship between X and Y is linear.

To illustrate the use of Table 26.2, let's consider the correlation we have just calculated for 5-year weights of cows and their female offspring. In this case, we obtained a value of 0.237 for r, and the value of N was 10. Under these conditions, *df* equals 10 (the number of paired observations) minus 2, which equals 8. To be valid at a 5% significance level, the value of r would have to

TABLE **26.2**

Values of *r* at the 5% and 1% Significance Levels

Degrees of Freedom (*df*)	5%	1%	Degrees of Freedom (*df*)	5%	1%
1	0.997	1.000	24	0.388	0.496
2	0.950	0.990	25	0.381	0.487
3	0.878	0.959	26	0.374	0.478
4	0.811	0.917	27	0.367	0.470
5	0.754	0.874	28	0.361	0.463
6	0.707	0.834	29	0.355	0.456
7	0.666	0.798	30	0.349	0.449
8	0.632	0.765	35	0.325	0.418
9	0.602	0.735	40	0.304	0.393
10	0.576	0.708	45	0.288	0.372
11	0.553	0.684	50	0.273	0.354
12	0.532	0.661	60	0.250	0.325
13	0.514	0.641	70	0.232	0.302
14	0.497	0.623	80	0.217	0.283
15	0.482	0.606	90	0.205	0.267
16	0.468	0.590	100	0.195	0.254
17	0.456	0.575	125	0.174	0.228
18	0.444	0.561	150	0.159	0.208
19	0.433	0.549	200	0.138	0.181
20	0.423	0.537	300	0.113	0.148
21	0.413	0.526	400	0.098	0.128
22	0.404	0.515	500	0.088	0.115
23	0.396	0.505	1000	0.062	0.081

Note: *df* equals $N - 2$.
From J. T. Spence, B. J. Underwood (1976). *Elementary Statistics.* Prentice-Hall, Englewood Cliffs, New Jersey.

be 0.632 or higher. Because the value we obtained is much less than this, it is fairly likely that this value could have occurred as a matter of random sampling error. In this case, we cannot reject the null hypothesis, and, therefore, we cannot conclude the positive correlation is due to a true association between the weights of mothers and offspring.

In an actual experiment, however, a researcher would examine many more pairs of cows and offspring, perhaps 500 to 1000. If a correlation of 0.237 was observed for $N = 1000$, the value would be significant at the 1% level. In this case, we would reject the null hypothesis that weights are not associated with each other. Instead, we would conclude that a real association occurs between the weights of mothers and their offspring. In fact, these kinds of experiments have been done for cattle weights, and the correlations between mothers and offspring have often been found to be significant.

If a statistically significant correlation is obtained, how do we interpret its meaning? An r value that is statistically significant suggests a true association, but it does not necessarily imply a cause-and-effect relationship. When parents and offspring display a significant correlation for a trait, we should not jump to the conclusion that genetics is the underlying cause of the positive association. In many cases, parents and offspring share similar environments, so the positive association might be rooted in environmental factors. In general, correlations are quite useful in identifying positive or negative associations between two variables. We should use caution, however, because this statistic, by itself, cannot prove that the association is due to cause and effect.

A **regression analysis** may be used when researchers suspect, or when their experimentation has shown, that two variables are related due to cause and effect—that one variable (the independent variable) affects the outcome of another (the dependent variable). Researchers use a regression analysis to predict how much the dependent variable changes in response to the independent variable. This approach is described in Solved Problem S4 at the end of the chapter.

26.2 REVIEWING THE KEY CONCEPTS

- Statistical methods are used to analyze quantitative traits. These include the mean, variance, standard deviation, covariance, and correlation (see Figure 26.2, Table 26.2).

26.2 COMPREHENSION QUESTIONS

1. The variance is
   a. a measure of the variation around the mean.
   b. computed as a squared deviation.
   c. higher when there is less phenotypic variation.
   d. both a and b are correct.
2. Which of the following statistics is used to compare two variables?
   a. Mean
   b. Correlation
   c. Variance
   d. Standard deviation

26.3 POLYGENIC INHERITANCE

Thus far, we have seen that quantitative traits tend to show a continuum of variation and can be analyzed with various statistical tools. At the beginning of the 1900s, a great debate focused on the inheritance of quantitative traits. The biometric school, founded by Francis Galton and Karl Pearson, argued that these types of traits are not controlled by discrete genes that affect phenotypes in a predictable way. To some extent, the biometric school favored the idea of blending inheritance, which proposed that the traits of each parent would mix together to form the traits of the offspring (see Chapter 3).

Alternatively, the followers of Mendel, led by William Bateson in England and William Castle in the United States, held firmly to the idea that traits are governed by genes, which are inherited as discrete units. As we now know, Bateson and Castle were correct in their assertion that genes are inherited as discrete units. However, as we will see in this section, studying quantitative traits is difficult because these traits are controlled by multiple genes and substantially influenced by environmental factors.

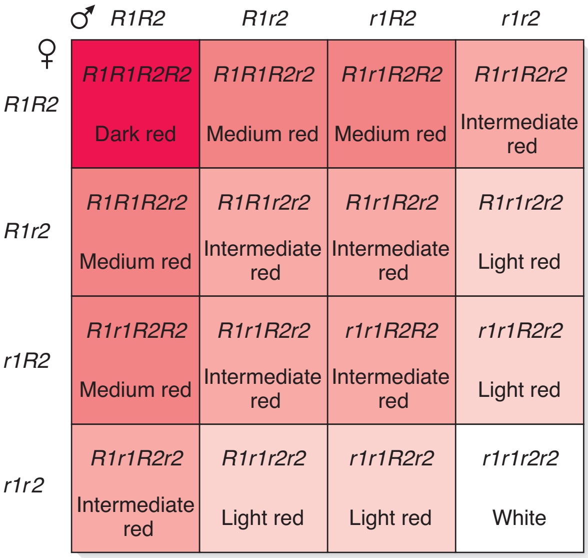
Most quantitative traits are polygenic and exhibit a continuum of phenotypic variation. The term **polygenic inheritance** refers to the transmission of a trait governed by two or more different genes. The location on a chromosome that harbors one or more genes that affect the outcome of a quantitative trait is called a **quantitative trait locus (QTL).** As discussed later, QTLs are chromosomal regions that are identified by genetic mapping. Because such mapping usually locates the QTL to a relatively large chromosomal region, a QTL may contain a single gene or two or more closely linked genes that affect a quantitative trait.

Just a few years ago, it was extremely difficult for geneticists to determine the inheritance patterns for genes underlying polygenic traits, particularly those determined by three or more genes having multiple alleles for each gene. Recently, however, molecular genetic tools (described in Chapters 21 and 23) have greatly enhanced our ability to find regions in the genome where QTLs are likely to reside. This has been a particularly exciting advance in the field of quantitative genetics. In some cases, the identification of QTLs may allow for the improvement of quantitative traits in agriculturally important species.

Polygenic Inheritance and Environmental Factors May Produce a Continuum of Phenotypes

The first experiment demonstrating that continuous variation is related to polygenic inheritance was conducted by the Swedish geneticist Herman Nilsson-Ehle in 1909. He studied the inheritance of red pigment in the hull of bread wheat, *Triticum aestivum* (**Figure 26.3a**). When true-breeding plants with white hulls were crossed to a variety with red hulls, the F_1 generation had an intermediate color. When the F_1 generation was allowed

(a) Red and white hulls of wheat



♀ \ ♂	*R1R2*	*R1r2*	*r1R2*	*r1r2*
R1R2	*R1R1R2R2* Dark red	*R1R1R2r2* Medium red	*R1r1R2R2* Medium red	*R1r1R2r2* Intermediate red
R1r2	*R1R1R2r2* Medium red	*R1R1r2r2* Intermediate red	*R1r1R2r2* Intermediate red	*R1r1r2r2* Light red
r1R2	*R1r1R2R2* Medium red	*R1r1R2r2* Intermediate red	*r1r1R2R2* Intermediate red	*r1r1R2r2* Light red
r1r2	*R1r1R2r2* Intermediate red	*R1r1r2r2* Light red	*r1r1R2r2* Light red	*r1r1r2r2* White

(b) *R1r1R2r2* x *R1r1R2r2*

FIGURE 26.3 **The Nilsson-Ehle experiment studying how continuous variation is related to polygenic inheritance in wheat.** **(a)** Red (top) and white (bottom) varieties of wheat, *Triticum aestivum*. **(b)** Nilsson-Ehle carefully categorized the colors of the hulls in the F_2 generation and discovered that they followed a 1:4:6:4:1 ratio. This occurs because the contributions of the red alleles are additive.

Genes→Traits In this example, two genes, with two alleles each (red and white), govern hull color. Offspring can display a range of colors, depending on how many copies of the red allele they inherit. If an offspring is homozygous for the red allele of both genes, it will have very dark red hulls. By comparison, if it carries three red alleles and one white allele, it will be medium red (which is not quite as deep in color). In this way, this polygenic trait can exhibit a range of phenotypes from dark red to white.

Concept check: *When we say these alleles are additive, what does this mean?*

to self-fertilize, great variation in redness was observed in the F_2 generation, ranging from white, light red, intermediate red, medium red, and dark red. An unsuspecting observer might conclude that this F_2 generation displayed a continuous variation in hull color. However, as shown in **Figure 26.3b**, Nilsson-Ehle carefully categorized the colors of the hulls and discovered that they followed a 1:4:6:4:1 ratio. He concluded that this species has two different genes that control hull color, each gene existing in a red or white allelic form. He hypothesized that these two loci must contribute additively to the color of the hull. In this case, each red allele carried by a given plant contributed to the red color of the hull. Plants carrying more red alleles had a deeper red color.

Later, researchers discovered a third gene that also affects hull color. The two strains that Nilsson-Ehle had used in his original experiments must have been homozygous for the white allele of this third gene. It makes sense that wheat would have two copies of three genes that affect hull color because we now know that *T. aestivum* is a hexaploid derived from three closely related diploid species, as discussed in Chapter 8. Therefore, *T. aestivum* has six copies of many genes.

As we have just seen, Nilsson-Ehle categorized wheat hull colors into several discrete genotypic categories. However, for many polygenic traits, this is difficult or impossible. In general, as the number of genes controlling a trait increases and the influence of the environment increases, the categorization of phenotypes into discrete genotypic classes becomes increasingly difficult, if not impossible. Therefore, a Punnett square cannot be used to analyze most quantitative traits. Instead, statistical methods, which are described later, must be employed.

Figure 26.4 illustrates how genotypes and phenotypes may overlap for polygenic traits. In this example, the environment (sunlight, soil conditions, and so forth) may affect the phenotypic outcome of a trait in plants (namely, seed weight). Figure 26.4a considers a situation in which seed weight is controlled by one gene with light (*w*) and heavy (*W*) alleles. A heterozygous plant (*Ww*) is allowed to self-fertilize. When the weight is only slightly influenced by variation in the environment, as seen on the left, the light, intermediate, and heavy seeds fall into separate, well-defined categories. When the environmental variation has a greater effect on seed weight, as shown on the right, more phenotypic variation is found in seed weight within each genotypic class. The variance in the frequency distribution on the right is much higher. Even so, most individuals can still be classified into the three main categories.

By comparison, Figure 26.4b illustrates a situation in which seed weight is governed by three genes instead of one, each existing in light and heavy alleles. When the environmental variation is low and/or plays a minor role in the outcome of this trait, a cross between two heterozygotes is expected to produce a 1:6:15:20:15:6:1 ratio. As shown in the upper illustration in Figure 26.4b, nearly all individuals fall within a phenotypic category that corresponds to their genotype. When the environment has a more variable effect on phenotype, as shown in the lower illustration, the situation becomes more ambiguous. For example, individuals with one *w* allele and five *W* alleles have a phenotype that overlaps with that of individuals having six *W* alleles or two *w* alleles and four *W* alleles. Therefore, it becomes difficult to categorize each phenotype into a unique genotypic class. Instead, the trait displays a continuum ranging from light to heavy seed weight.

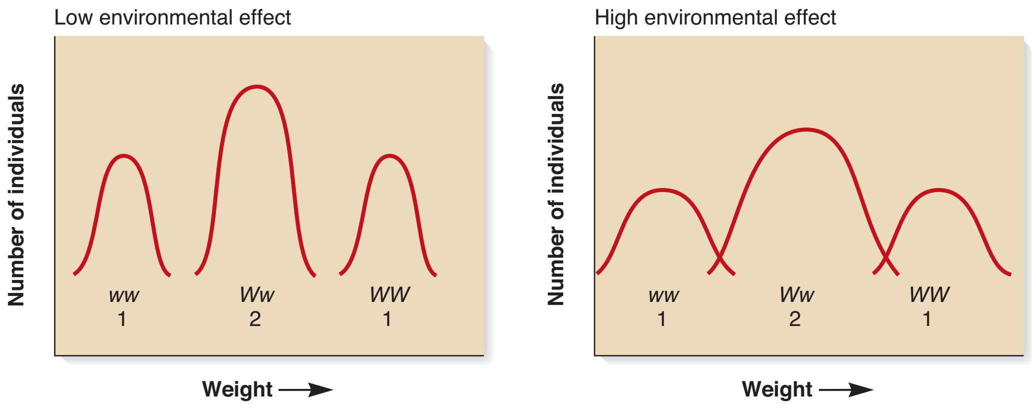
Quantitative Trait Loci (QTLs) Are Now Mapped by Linkage to Known Molecular Markers

As discussed in Chapters 21 and 23, newer research techniques have identified molecular markers, such as restriction fragment length polymorphism (RFLPs) and microsatellites, which serve as reference points along chromosomes. These markers have been used to construct detailed genetic maps of several species' genomes. Once a genome map is obtained, it becomes much easier to determine the locations of genes that affect a quantitative trait. In addition to model organisms such as *Drosophila*, *Arabidopsis*, *Caenorhabditis elegans*, and mice, detailed molecular maps have been obtained for many species of agricultural importance. These include crops such as corn, rice, and tomatoes, as well as livestock such as cattle, pigs, sheep, and chickens.

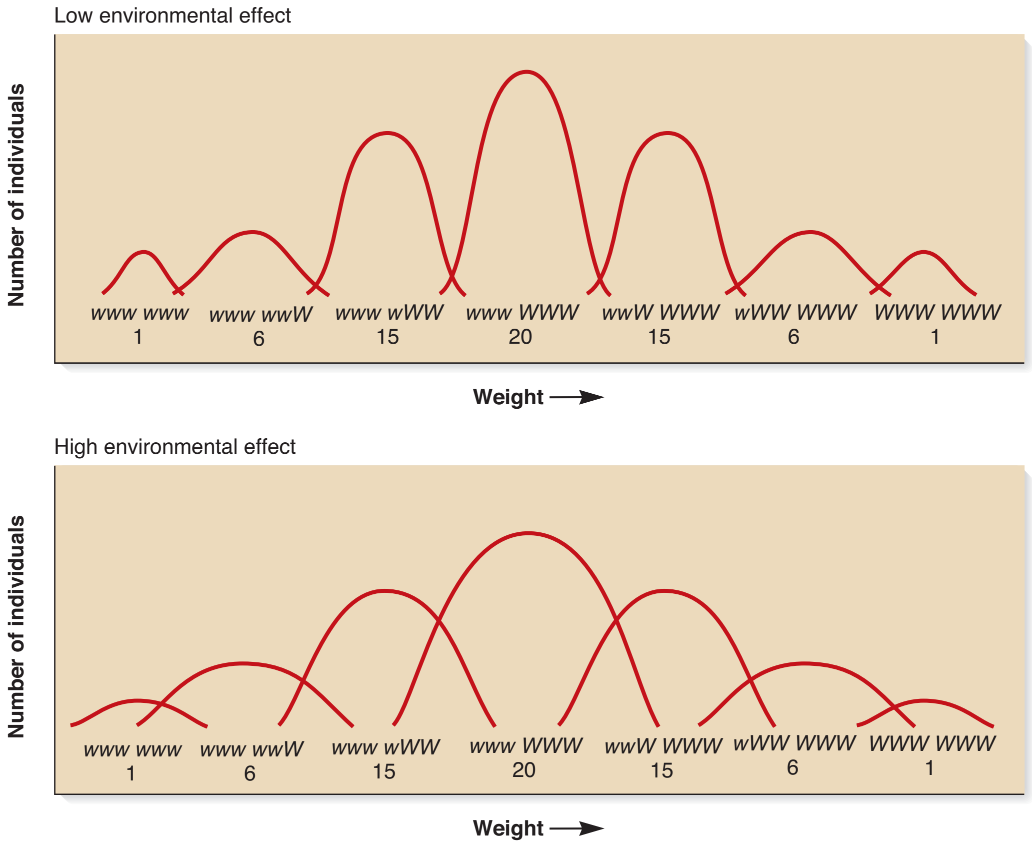
To map the genes in a eukaryotic species, researchers now determine their locations by identifying their linkage to molecular markers. This approach is described in Chapter 23 (see Figures 23.5–23.7). In 1989, Eric Lander and David Botstein extended this technique to identify QTLs that govern a quantitative trait. The basis of **QTL mapping** is the association between genetically determined phenotypes for quantitative traits and molecular markers such as RFLPs, microsatellites, and single-nucleotide polymorphisms (SNPs). In this approach, the sites of many molecular markers are already known, and the researcher identifies QTLs that are close to particular markers.

The general strategy for QTL mapping is shown in **Figure 26.5**. This figure depicts two different strains of a diploid species with four chromosomes per set. The strains are highly inbred, which means they are homozygous for most molecular markers and genes. They differ in two important ways. First, the two strains differ with regard to many molecular markers; the sites of many markers along each chromosome are already known. These markers are designated 1A and 1B, 2A and 2B, and so forth. The markers 1A and 1B mark the same chromosomal location in this species, namely, the upper tip of chromosome 1. However, the two markers are distinguishable in the two strains at the molecular level. For example, 1A might be a microsatellite that is 148-bp, whereas 1B might be 212-bp. Second, the two strains differ in a quantitative trait of interest. In this example, the strain on the left produces large fruit, whereas the strain on the right produces small fruit. The unknown genes affecting this trait are designated with the letter X. A black X indicates a QTL that harbors alleles that promote large fruit, and a blue X is the same site that carries alleles that promote small fruit. Prior to conducting their crosses, researchers would not know the chromosomal locations of the QTLs shown in this figure. The purpose of the experiment is to determine their locations.

With these ideas in mind, the protocol shown in Figure 26.5 begins by mating the two inbred strains to each other and then backcrossing the F_1 offspring to both parental strains. This produces an F_2 generation with a great degree of variation. The



(a) ***Ww* x *Ww***



(b) ***WWW www* x *WWW www***

FIGURE 26.4 **How genotypes and phenotypes may overlap for polygenic traits.** **(a)** Situations in which seed weight is controlled by one gene, existing in light (*w*) and heavy (*W*) alleles. **(b)** Situations in which seed weight is governed by three genes instead of one, each existing in light and heavy alleles. Note: The 1:2:1 and 1:6:15:20:15:6:1 ratios were derived by using a Punnett square and assuming a cross between individuals that are both heterozygous for three different genes.

Genes→Traits The ability of geneticists to correlate genotype and phenotype depends on how many genes are involved and how much the environment causes the phenotype to vary. In (a), a single gene influences weight. In the graph on the left side, the environment does not cause much variation in weight. This makes it easy to distinguish the three genotypes. There is no overlap in the weights of *ww*, *Ww*, and *WW* individuals. In the graph on the right side, the environment causes more variation in weight. In this case, a few individuals with *ww* genotypes have the same weight as a few individuals with *Ww* genotypes; and a few *Ww* genotypes have the same weight as *WW* genotypes. As shown in (b), it becomes even more difficult to distinguish genotype based on phenotype when three genes are involved. The overlaps are minor when the environment does not cause much weight variation. However, when the environment causes substantial phenotypic variation, the overlaps between genotypes and phenotypes are very pronounced and greatly confound genetic analysis.

Concept check: *Explain how gene number and environmental variation affects the overlaps between phenotypes and different genotypes.*

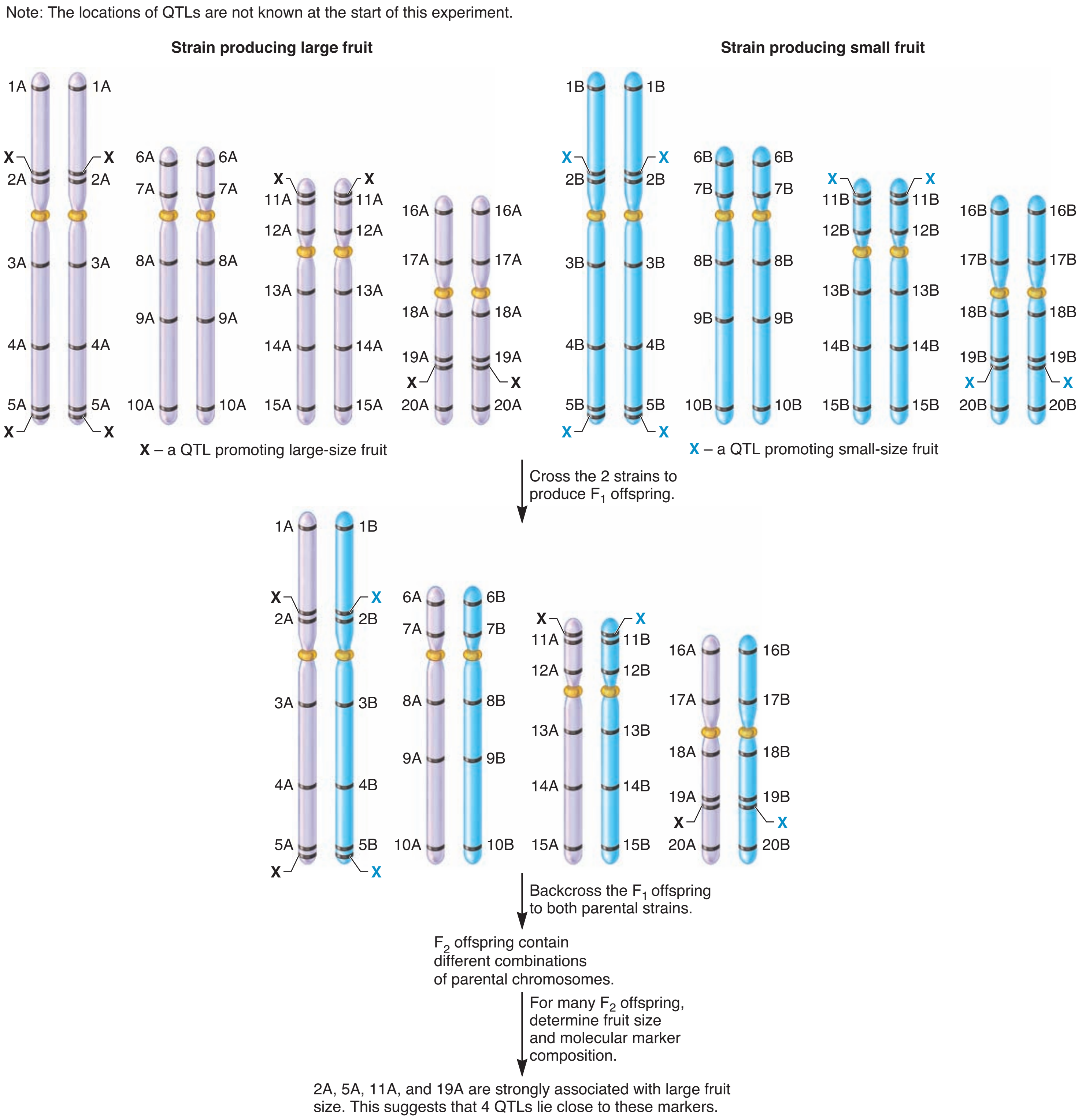

Note: The locations of QTLs are not known at the start of this experiment.
Strain producing large fruit
Strain producing small fruit
1A
X
2A
3A
4A
5A
6A
7A
8A
9A
10A
11A
12A
13A
14A
15A
16A
17A
18A
19A
20A
1B
2B
3B
4B
5B
6B
7B
8B
9B
10B
11B
12B
13B
14B
15B
16B
17B
18B
19B
20B
X – a QTL promoting large-size fruit
X – a QTL promoting small-size fruit
Cross the 2 strains to produce F_1 offspring.
Backcross the F_1 offspring to both parental strains.
F_2 offspring contain different combinations of parental chromosomes.
For many F_2 offspring, determine fruit size and molecular marker composition.
2A, 5A, 11A, and 19A are strongly associated with large fruit size. This suggests that 4 QTLs lie close to these markers.

FIGURE 26.5 **The general strategy for QTL mapping via molecular markers.** Two different inbred strains have four chromosomes per set. The strain on the left produces large fruit, and the strain on the right produces small fruit. The goal of this mapping strategy is to locate the unknown genes affecting this trait, which are designated with the letter X. A black X indicates a site promoting large fruit, and a blue X is a site promoting small fruit. The two strains differ with regard to many molecular markers designated 1A and 1B, 2A and 2B, and so forth. The two strains are mated, and then the F_1 offspring are backcrossed to the parental strains. Many F_2 offspring are then examined for their fruit size and to determine which molecular markers are found in their chromosomes. The data are analyzed by computer programs that can statistically associate the phenotype (e.g., fruit size) with particular markers. Markers found throughout the genome of this species provide a way to locate several different genes that may affect the outcome of a single quantitative trait. In this case, the analysis predicts four QTLs promoting heavier fruit weight that are linked to regions of the chromosomes containing the following markers: 2A, 5A, 11A, and 19A.

Concept check: *What are the two ways that strains A and B differ?*

F_2 offspring are then characterized in two ways. First, they are examined for their fruit size, and second, a sample of cells is analyzed to determine which molecular markers are found in their chromosomes. The goal is to find an association between particular molecular markers and fruit size. For example, 2A is strongly associated with large size, whereas 2B is strongly associated with small size. By comparison, 9A and 9B are not associated with large or small size, because a QTL affecting this trait is not found on this chromosome. Also, markers such as 14A and 14B, which are fairly far away from a QTL, are not strongly associated with a particular QTL. Markers that are on the same chromosome but far away from a QTL are often separated from the QTL during meiosis in the F_1 heterozygote due to crossing over. Only closely linked markers are strongly associated with a particular QTL.

Overall, QTL mapping involves the analysis of a large number of markers and offspring. The data are analyzed by computer programs that can statistically associate the phenotype (e.g., fruit size) with particular markers. Markers found throughout the genome of a species provide a way to identify the locations of several different genes that possess allelic differences that may affect the outcome of a quantitative trait.

As an example of QTL mapping, in 1988, Andrew Paterson and his colleagues examined quantitative trait inheritance in the tomato. They studied a domestic strain of tomato and a South American green-fruited variety. These two strains differed in their RFLPs, and they also exhibited dramatic differences in three agriculturally important characteristics: fruit mass, soluble solids content, and fruit pH. The researchers crossed the two strains and then backcrossed the offspring to the domestic tomato. A total of 237 plants were then examined with regard to 70 known RFLP markers. In addition, between 5 and 20 tomatoes from each plant were analyzed with regard to fruit mass, soluble solids content, and fruit pH. Using this approach, the researchers were able to map genes contributing much of the variation in these traits to particular sites along the tomato chromosomes. They identified six loci causing variation in fruit mass, four affecting soluble solids content, and five with effects on fruit pH.

More recently, the DNA sequence of the entire genome of many species has been determined. In such cases, the mapping of QTLs to a defined chromosomal region may allow researchers to analyze the DNA sequence in that region and to identify one or more genes that influence the trait of interest.

26.3 REVIEWING THE KEY CONCEPTS

- Polygenic inheritance refers to an inheritance pattern in which multiple genes affect a single trait.
- The locations along a chromosome that contain genes affecting a quantitative trait are called quantitative trait loci (QTLs).
- Polygenic inheritance and environmental factors may cause a quantitative trait to fall along a continuum (see Figures 26.3, 26.4).
- Quantitative trait loci are mapped by their linkage to molecular markers (see Figure 26.5).

26.3 COMPREHENSION QUESTIONS

1. For quantitative traits, genotypes and phenotypes tend to overlap because
    a. the trait changes over time.
    b. they are polygenic.
    c. environmental variation affects the trait.
    d. both b and c are correct.
2. A QTL is a __________ where one or more genes affecting a quantitative trait are ________ .
    a. site in a cell, located
    b. site in a chromosome, located
    c. site in a cell, expressed
    d. site in a chromosome, expressed
3. To map QTLs, strains are crossed that differ with regard to
    a. a quantitative trait.
    b. molecular markers.
    c. a quantitative trait and molecular markers.
    d. a quantitative trait and a discontinuous trait.

26.4 HERITABILITY

As we have just seen, recent approaches in molecular mapping have enabled researchers to identify the genes that contribute to a quantitative trait. The other key factor that affects the phenotypic outcomes of quantitative traits is the environment. All traits of organisms are influenced by both genetics and the environment, and this is particularly pertinent in the study of quantitative traits. Researchers want to understand how variation, both genetic and environmental, affects the phenotypic results.

The term **heritability** refers to the amount of phenotypic variation within a group of individuals that is due to genetic variation. Genes play a role in the development of essentially all of an organism's traits. Even so, variation of a trait in a population may be due entirely to environmental variation, entirely to genetic variation, or to a combination of the two. If all of the phenotypic variation in a group is due to genetic variation, the heritability would have a value of 1. If all of the variation is due to environmental effects, the heritability would equal 0. For most groups of organisms, the heritability for a given trait lies between these two extremes. For example, both genes and diet affect the size that an individual will attain. Some individuals inherit alleles that tend to make them large, and a proper diet also promotes larger size. Other individuals inherit alleles that make them small, and an inadequate diet may contribute to small size. Taken together, both genetics and the environment influence the phenotypic results.

In the study of quantitative traits, a primary goal is to determine how much of the phenotypic variation arises from genetic variation and how much comes from environmental variation. In this section, we examine how geneticists analyze the genetic and environmental components that affect quantitative traits.

Both Genetic Variance and Environmental Variance May Contribute to Phenotypic Variance

Earlier, we examined the amount of phenotypic variation within a group by calculating the variance. Geneticists partition quantitative trait variation into components that are attributable to the following different causes:

Genetic variation (V_G)
Environmental variation (V_E)
Variation due to interactions between genetic and environmental factors ($V_{G \times E}$)
Variation due to associations between genetic and environmental factors ($V_{G \leftrightarrow E}$)

Let's begin by considering a simple situation in which V_G and V_E are the only factors that determine phenotypic variance, and the genetic and environmental factors are independent of each other. If so, then the total variance for a trait in a group of individuals is

$$V_T = V_G + V_E$$

where

V_T	is the total variance. It reflects the amount of variation that is measured at the phenotypic level.
V_G	is the relative amount of variance due to genetic variation.
V_E	is the relative amount of variance due to environmental variation.

Why is this equation useful? The partitioning of variance into genetic and environmental components allows us to estimate their relative importance in influencing the variation within a group. If V_G is very high and V_E is very low, genetics plays a greater role in promoting variation within a group. Alternatively, if V_G is low and V_E is high, environmental factors underlie much of the phenotypic variation. As described later in this chapter, a livestock breeder might want to apply selective breeding if V_G for an important (quantitative) trait is high. In this way, the characteristics of the herd may be improved. Alternatively, if V_G is negligible, it would make more sense to investigate and manipulate the environmental causes of phenotypic variation.

With experimental and domesticated species, one possible way to determine V_G and V_E is by comparing the variation in traits between genetically identical and genetically disparate groups. For example, researchers have followed the practice of **inbreeding** to develop genetically homogeneous strains of mice. Inbreeding in mice involves many generations of brother-sister matings, which eventually produces strains that are **monomorphic** for all or nearly all of their genes. The term *monomorphic* means that all members of a population are homozygous for the same allele of a given gene. Within such an inbred strain of mice, V_G equals zero. Therefore, all phenotypic variation is due to V_E. When studying quantitative traits such as weight, an experimenter might want to know the genetic and environmental variance for a different, genetically heterogeneous group of mice. To do so, the genetically homogeneous and heterogeneous mice could be raised under the same environmental conditions and their weights measured. The phenotypic variance for weight could then be calculated as described earlier. Let's suppose we obtained the following results:

$V_T = 15\ g^2$ for the group of genetically homogeneous mice

$V_T = 22\ g^2$ for the group of genetically heterogeneous mice

In the case of the homogeneous mice, $V_T = V_E$, because V_G equals 0. Therefore, V_E equals 15 g^2. To estimate V_G for the heterogeneous group of mice, we could assume that V_E (i.e., the environmentally produced variance) is the same for them as it is for the homogeneous mice, because the two groups were raised in identical environments. This assumption allows us to calculate the genetic variance for the heterogeneous mice.

$$V_T = V_G + V_E$$

$$22\ g^2 = V_G + 15\ g^2$$

$$V_G = 7\ g^2$$

This result tells us that some of the phenotypic variance in the genetically heterogeneous group is due to the environment (namely, 15 g^2) and some (7 g^2) is due to genetic variation in alleles that affect weight.

Phenotypic Variation May Also Be Influenced by Interactions and Associations Between Genotype and the Environment

Thus far, we have considered the simple situation in which genetic variation and environmental variation are independent of each other and affect the phenotypic variation in an additive way. As another example, let's suppose that three genotypes, *HH*, *Hh*, and *hh*, affect height, producing tall, intermediate, and short plants, respectively. Greater sunlight makes the plants grow taller regardless of their genotypes. In this case, our assumption that $V_T = V_G + V_E$ would be reasonably valid.

However, let's consider a different environmental factor such as minerals in the soil. As a hypothetical example, let's suppose the *h* allele decreases the function of a protein involved with mineral uptake from the soil. In this case, the *Hh* and *hh* plants are shorter because they cannot take up a sufficient supply of certain minerals to support maximal growth, whereas the *HH* plants are not limited by mineral uptake. According to this hypothetical scenario, adding minerals to the soil enhances the growth rate of *hh* plants by a large amount and the *Hh* plants by a smaller amount, because the minerals are more easily taken up by the plants (**Figure 26.6**). The height of *HH* plants is not affected by mineral supplementation. When the environmental effects on phenotype differ according to genotype, this phenomenon is called a **genotype-environment interaction.** Variation due to interactions between genetic and environmental factors is termed $V_{G \times E}$, as noted earlier.

Interactions between genetic and environmental factors are common. As an example, **Table 26.3** shows results from a study conducted in 2000 by Cristina Vieira, Trudy Mackay, and colleagues in which they investigated genotype-environment interaction for QTLs affecting life span in *Drosophila melanogaster*. The data seen in the table compare the life span in days of male and female flies from two different strains of *D. melanogaster* raised at different temperatures. Because males and females differ in their sex chromosomes and gene expression patterns, they can be viewed as having different genotypes. The effects of

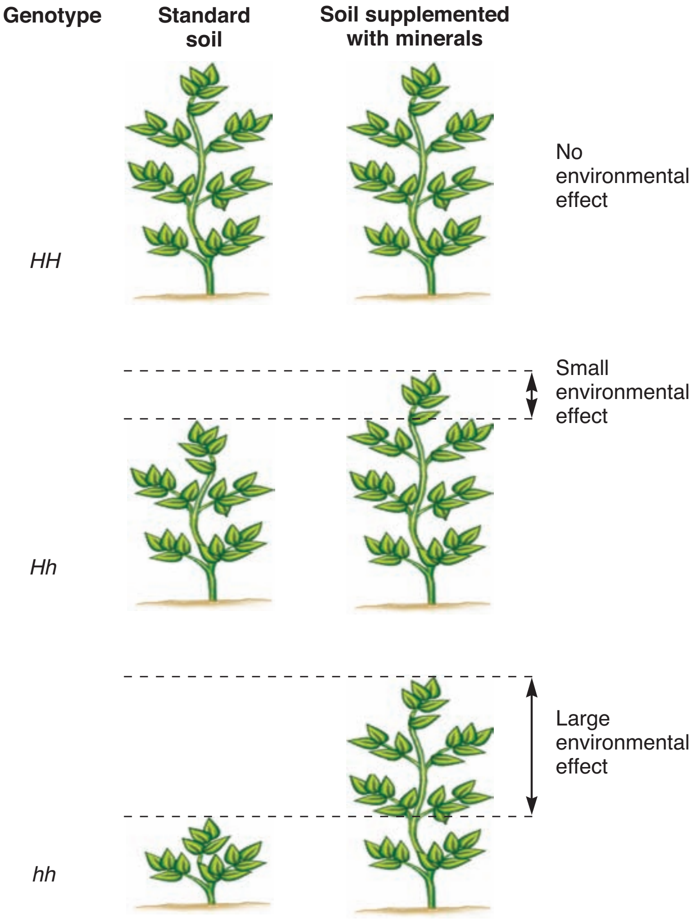

Genotype
Standard soil
Soil supplemented with minerals
No environmental effect
HH
Small environmental effect
Hh
Large environmental effect
hh

FIGURE 26.6 A schematic example of genotype-environment interaction. When grown in standard soil, the three genotypes *HH*, *Hh*, and *hh* show tall, intermediate, and short heights, respectively. When the soil is supplemented with minerals, a large effect is seen on the *hh* genotype and a smaller effect on the *Hh* genotype. The *HH* genotype is unaffected by the environmental change.

TABLE **26.3**

Longevity of Two Strains of *Drosophila melanogaster**

	Strain A		Strain B	
Temperature	Male	Female	Male	Female
Standard	33.6	39.5	37.5	28.9
High	36.3	33.9	23.2	28.6
Low	77.5	48.3	45.8	77.0

*Longevity was measured in the mean number of days of survival. Strains A and B were inbred strains of *D. melanogaster* called Oregon and 2b, respectively. The standard-, high-, and low-temperature conditions were 25˚C, 29˚C, and 14˚C, respectively.

environmental changes depended greatly on the strain and the sex of the flies. Under standard culture conditions, the females of strain A had the longest life span, whereas females of strain B had the shortest. In strain A, high temperature increased the longevity of males and decreased the longevity of females. In contrast, under hotter conditions, the longevity of males of strain B was dramatically reduced, whereas females of this same strain were not significantly affected. Lower growth temperature also had different effects in these two strains. Although low temperature increased the longevity of both strains, the effects were most dramatic in the males of strain A and the females of strain B. Taken together, these results illustrate the potential complexity of genotype-environmental interaction when measuring a quantitative trait such as life span.

Another issue confronting geneticists is that genotypes may not be randomly distributed in all possible environments. When certain genotypes are preferentially found in particular environments, this phenomenon is called a **genotype-environment association** ($V_{G\leftrightarrow E}$). When such an association occurs, the effects of genotype and environment are not independent of each other, and the association needs to be considered when determining the effects of genetic and environmental variation on the total phenotypic variation. Genotype-environment associations are very common in the study of human genetics, in which large families tend to have more similar environments than the population as a whole. One way to evaluate this effect is to compare different genetic relationships, such as identical versus fraternal twins. We will examine this approach later in the chapter. Another strategy that geneticists might follow is to analyze siblings that have been adopted by different parents at birth. Their environmental conditions tend to be more disparate, and this may help to minimize the effects of genotype-environment association.

Heritability Is the Relative Amount of Phenotypic Variation That Is Due to Genetic Variation

Another way to view variance is to focus our attention on the genetic contribution to phenotypic variation. Heritability is the proportion of the phenotypic variance that is attributable to genetic variation. If we assume again that environment and genetics are independent and the only two factors affecting phenotype, then

$$h_B^2 = V_G/V_T$$

where

h_B^2 is the heritability in the broad sense
V_G is the variance due to genetics
V_T is the total phenotypic variance, which equals $V_G + V_E$

The heritability defined here, h_B^2, called **broad-sense heritability,** takes into account different types of genetic variation that may affect the phenotype. As we have seen throughout this textbook, genes can affect phenotypes in various ways. As described earlier, the Nilsson-Ehle experiment showed that the alleles determining hull color in wheat affect the phenotype in an additive way. A heterozygote shows a phenotype that is intermediate between the respective homozygotes. Alternatively, alleles affecting other traits may show a dominant/recessive relationship. In this case, the alleles are not strictly additive, because the heterozygote has a phenotype closer to, or perhaps the same as, the homozygote containing two copies of the dominant allele.

For example, Mendel discovered that both *PP* and *Pp* pea plants have purple flowers. In addition, another complicating factor is epistasis (described in Chapter 5), in which the alleles for one gene can mask the phenotypic expression of the alleles of another gene. To account for these differences, geneticists usually subdivide V_G into these three different genetic categories:

$$V_G = V_A + V_D + V_I$$

where

V_A is the variance due to the additive effects of alleles.

V_D is the variance due to the effects of alleles that follow a dominant/recessive pattern of inheritance.

V_I is the variance due to the effects of alleles that interact in an epistatic manner.

In analyzing quantitative traits, geneticists may focus on V_A and ignore the contributions of V_D and V_I. They do this for scientific as well as practical reasons. For some quantitative traits, the additive effects of alleles may play a primary role in the phenotypic outcome. In addition, when the alleles behave additively, we can predict the outcomes of crosses based on the quantitative characteristics of the parents. The heritability of a trait due to the additive effects of alleles is called **narrow-sense heritability:**

$$h_N^2 = V_A/V_T$$

For many quantitative traits, the value of V_A may be relatively large compared with V_D and V_I. In such cases, the determination of the narrow-sense heritability provides an estimate of the broad-sense heritability.

How can the narrow-sense heritability be determined? In this chapter, we will consider two common ways. As discussed later, one way to calculate the narrow-sense heritability involves selective breeding practices, which are done with agricultural species. A second common strategy to determine h_N^2 involves the measurement of a quantitative trait among groups of genetically related individuals. For example, agriculturally important traits, such as egg weight in poultry, can be analyzed in this way. To calculate the heritability, a researcher determines the observed egg weights between individuals whose genetic relationships are known, such as a mother and her female offspring. These data can then be used to compute a correlation between the parent and offspring, using the methods described earlier. The narrow-sense heritability is then calculated as

$$h_N^2 = r_{obs}/r_{exp}$$

where

r_{obs} is the observed phenotypic correlation between related individuals

r_{exp} is the expected correlation based on the known genetic relationship

In our example, r_{obs} is the observed phenotypic correlation between parent and offspring. In particular research studies, the observed phenotypic correlation for egg weights between mothers and daughters has been found to be about 0.3 (although this varies among strains). The expected correlation, r_{exp}, is based on the known genetic relationship. A parent and child share 50% of their genetic material, so r_{exp} equals 0.50. So,

$$\begin{aligned} h_N^2 &= r_{obs}/r_{exp} \\ &= 0.3/0.50 \\ &= 0.60 \end{aligned}$$

(Note: For siblings, $r_{exp} = 0.50$; for identical twins, $r_{exp} = 1.0$; and for an aunt-niece relationship, $r_{exp} = 0.25$.)

According to this calculation, about 60% of the phenotypic variation in egg weight is due to additive genetic variation; the other 40% is due to the environment.

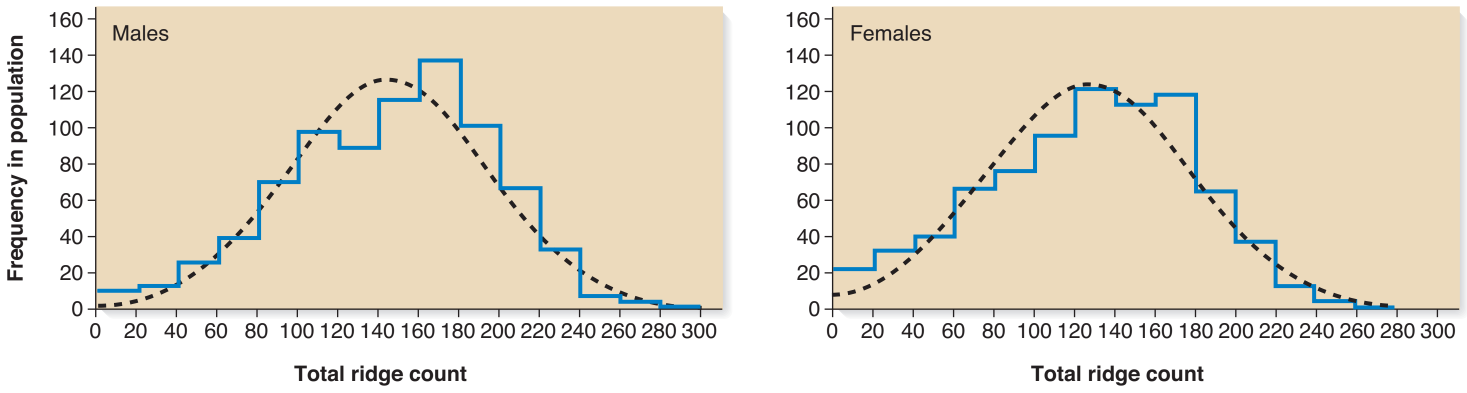
When calculating heritabilities from correlation coefficients, keep in mind that this computation assumes that genetics and the environment are independent variables. However, this is not always the case. The environments of parents and offspring are often more similar to each other than they are to those of unrelated individuals. As mentioned earlier, there are several ways to minimize this confounding factor. First, in human studies, one may analyze the heritabilities from correlations between adopted children and their biological parents. Alternatively, one can examine a variety of relationships (aunt-niece, identical twins versus fraternal twins, and so on) and see if the heritability values are roughly the same in all cases. This approach was applied in the study that is described next.

Heritability of Dermal Ridge Count in Human Fingerprints Is Very High

Fingerprints are inherited as a quantitative trait. It has long been known that identical twins have fingerprints that are very similar, whereas fraternal twins show considerably less agreement. Galton was the first researcher to study fingerprint patterns, but this trait became more amenable to genetic studies in the 1920s, when Kristine Bonnevie, a Norwegian geneticist, developed a method for counting the number of ridges within a human fingerprint.

As shown in **Figure 26.7**, human fingerprints can be categorized as having an arch, loop, or whorl, or a combination of these patterns. The primary difference among these patterns is the number of triple junctions, each known as a triradius (Figure 26.7b and c). At a triradius, a ridge emanates in three different directions. An arch has zero triradii, a loop has one, and a whorl has two. In Bonnevie's method of counting, a line is drawn from a triradius to the center of the fingerprint. The ridges that touch this line are then counted. (Note: The triradius ridge itself is not counted, and the last ridge is not counted if it forms the center of the fingerprint.) With this method, one can obtain a ridge count for all 10 fingers. Bonnevie conducted a study on a small population and found that ridge count correlations were relatively high in genetically related individuals.

Sarah Holt, who was also interested in the inheritance of this quantitative trait, carried out a more exhaustive study of ridge counts in a British population. As shown in the following two graphs (see top of next page), in groups of 825 males and 825 females, the ridge count on all 10 fingers varied from 0 to 300, with mean values of approximately 145 for males (SD = 51.1) and 127 for females (SD = 52.5).



Based on these results, Holt decided to conduct a more detailed analysis of ridge counts by examining the fingerprint patterns of a large group of people and their close relatives. In the experiment of **Figure 26.8**, the ridge counts for pairs of related individuals were determined by the method described in Figure 26.7. The correlation coefficients for ridge counts were then calculated among the pairs of related or unrelated individuals. To estimate the narrow-sense heritability, the observed correlations were then divided by the expected correlations based on the known genetic relationships.

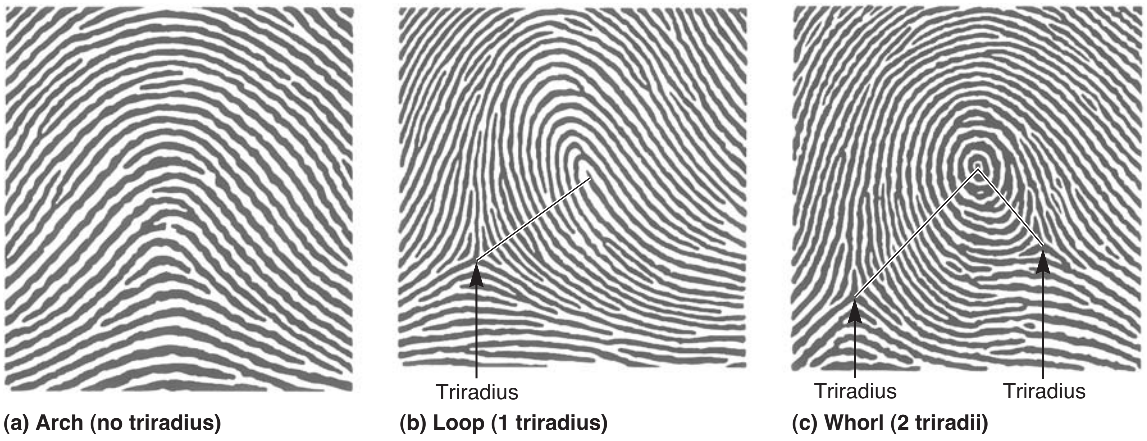


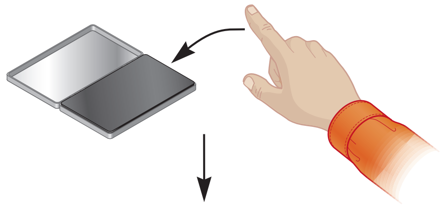
FIGURE 26.7 **Human fingerprints and the ridge count method of Bonnevie.** **(a)** This print has an arch rather than a triradius. The ridge count is zero. **(b)** This print has one triradius. A straight line is drawn from the triradius to the center of the print. The number of ridges dissecting this straight line is 13. **(c)** This print has two triradii. Straight lines are drawn from both triradii to the center. There are 16 ridges touching the left line and 7 touching the right line, giving a total ridge count of 23.

▶ THE HYPOTHESIS

Dermal ridge count has a genetic component. The goal of this experiment is to determine the contribution of genetics in the variation of dermal ridge counts.

▶ TESTING THE HYPOTHESIS — FIGURE 26.8 **Heritability of human fingerprint patterns.**

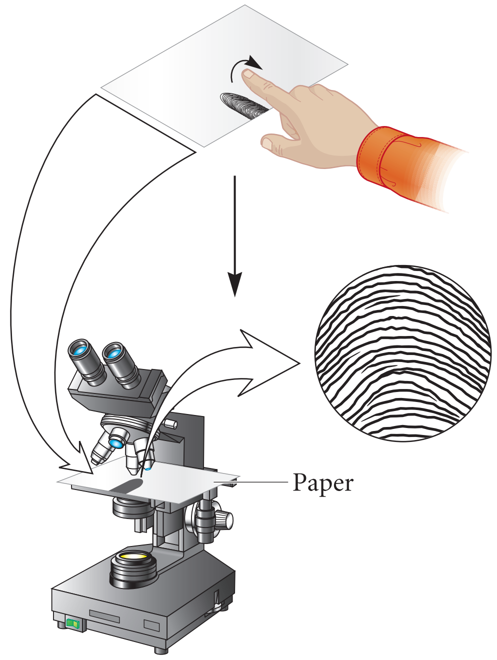
Starting material: A group of human subjects from Great Britain.

	Experimental level	Conceptual level
1. Take a person's finger and blot it onto an ink pad.		

2. Roll the person's finger onto a recording surface to obtain a print.

3. With a low-power binocular microscope, count the number of ridges, using the method described in Figure 26.7.

4. Calculate the correlation coefficients between different pairs of individuals as described earlier in this chapter.



See the data.

This is a method to measure a quantitative trait.

The correlation coefficient provides a way to determine the heritability for the quantitative trait.

▶ THE DATA

Type of Relationship	*Number of Pairs Examined*	*Correlation Coefficient* (r_{obs})	*Heritability* r_{obs}/r_{exp}
Parent-child	810	0.48 ± 0.04*	0.96
Parent-parent	200	0.05 ± 0.07	—[†]
Sibling-sibling	642	0.50 ± 0.04	1.00
Identical twins	80	0.95 ± 0.01	0.95
Fraternal twins	92	0.49 ± 0.08	0.98
			0.97 average heritability

*± = Standard error of the mean.

[†] Note: We cannot calculate a heritability value because the value for r_{exp} is not known. Nevertheless, the value for r_{obs} is very low, suggesting that there is a negligible correlation between unrelated individuals.

Adapted from S. B. Holt (1961). Quantitative genetics of fingerprint patterns. *Br Med Bull 17,* 247–250.

▶ INTERPRETING THE DATA

As seen in the data table, the results indicate that genetics plays the major role in explaining the variation in this trait. Genetically unrelated individuals (namely, parent-parent relationships) have a negligible correlation for this trait. By comparison, individuals who are genetically related have a substantially higher correlation. When the observed correlation coefficient is divided by the expected correlation coefficient based on the known genetic relationships, the average heritability value is 0.97, which is very close to 1.0.

What do these high heritability values mean? They indicate that nearly all of the variation in fingerprint pattern is due to genetic variation. Significantly, fraternal and identical twins have substantially different observed correlation coefficients, even though we expect they have been raised in very similar environments. These results support the idea that genetics is playing the major role in promoting variation and the results are not biased heavily by environmental similarities that may be associated with genetically related individuals. From an experimental viewpoint, the results show us how the determination of correlation coefficients between related individuals can provide insight regarding the relative contributions of genetics and environment to the variation of a quantitative trait.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookerconcepts.

Heritability Values Are Relevant Only to Particular Groups Raised in a Particular Environment

Table 26.4 describes heritability values that have been calculated for traits in particular populations. Unfortunately, heritability is a widely misunderstood concept. Heritability describes the amount of phenotypic variation due to genetic variation for a particular population raised in a particular environment. The words *variation*, *particular population*, and *particular environment* cannot be overemphasized. For example, in one population of cattle, the heritability for milk production may be 0.35, whereas in another group (with less genetic variation), the heritability may be 0.1. Second, if a group displays a heritability of 1.0 for a particular trait, this does not mean that the environment is unimportant in affecting the outcome of the trait. A heritability value of 1.0 only means that the amount of variation within this group is due to genetics. Perhaps the group has been raised in a relatively homogeneous environment, so the environment has not caused a significant amount of variation. Nevertheless, the environment may still be quite important. It just is not causing much variation within this particular group.

As a hypothetical example, let's suppose that we take a species of rodent and raise a group on a poor diet; we find their weights range from 1.5 to 2.5 pounds, with a mean weight of 2 pounds. We allow them to mate and then raise their offspring on a healthy diet of rodent chow. The weights of the offspring range from 2.5 to 3.5 pounds, with a mean weight of 3 pounds. In this hypothetical experiment, we might find a positive correlation in which the small parents tended to produce small offspring, and the large parents produce large offspring. The correlation of weights between parent and offspring might be, say, 0.5. In this case, the heritability for weight would be calculated as r_{obs}/r_{exp}, which equals 0.5/0.5, or 1.0. The value of 1.0 means that the total amount of phenotypic variation within this group is due to genetic variation among the individuals. The offspring vary from 2.5 to 3.5 pounds because of genetic variation, and also the parents range from 1.5 to 2.5 pounds because of genetics. However, as we see here, environment has played an important role. Presumably, the mean weight of the offspring is higher because of their better diet. This example is meant to emphasize the point that heritability tells us only the relative contributions of genetic variation and environment in influencing phenotypic *variation* in a *particular population* in a *particular environment*. Heritability does not describe the relative importance of these two factors in determining the outcomes of traits. When a heritability value is high, it does not mean that a change in the environment cannot have a major influence on the outcome of the trait.

TABLE **26.4**

Examples of Heritabilities for Quantitative Traits

Trait	Heritability Value*
Humans	
Stature	0.65
IQ testing ability	0.60
Cattle	
Body weight	0.65
Butterfat, %	0.40
Milk yield	0.35
Mice	
Tail length	0.40
Body weight	0.35
Litter size	0.20
Poultry	
Body weight	0.55
Egg weight	0.50
Egg production	0.10

*As emphasized in this chapter, these values apply to particular populations raised in particular environments. The value for IQ testing ability is an average value from many independent studies. The other values were taken from D. S. Falconer (1989). *Introduction to Quantitative Genetics*, 3rd ed. Longman, Essex, England.

With regard to the roles of genetics and environment (sometimes referred to as nature versus nurture), the topic of human intelligence has been hotly debated. As a trait, intelligence is difficult to define or to measure. Nevertheless, performance on an IQ test has been taken by some people as a reflection of intelligence ever since 1916, when the Stanford-Binet Intelligence Scale was first used in the United States. Even though such tests may have inherent bias and consider only a limited subset of human cognitive abilities, IQ tests still remain a method of assessing intelligence. By comparing IQ scores among related and unrelated individuals, various studies have attempted to estimate heritability values in selected human populations. These values have ranged from 0.3 to 0.8. A heritability value of around 0.6 is fairly common among many studies. Such a value indicates that over half of the heritability for IQ testing ability is due to genetic factors.

Let's consider what a value of 0.6 means, and what it does not mean. It means that 60% of the variation in IQ testing ability is due to genetic variation in a selected population raised in a particular environment. It does not mean that 60% of an individual's IQ testing ability is due to genetics and 40% is due to the environment. Heritability is meaningless at the level of a single individual. Furthermore, even at the population level, a heritability value of 0.6 does not mean that 60% of the IQ testing ability is due to genetics and 40% is due to the environment. Rather, it means that in the selected population that was examined, 60% of the variation in IQ testing ability is due to genetics, whereas 40% of the variation is due to the environment. Heritability is strictly a population value that pertains to variation.

26.4 REVIEWING THE KEY CONCEPTS

- Heritability is the amount of phenotypic variation within a group of individuals that is due to genetic variation.
- Genetic variance and environmental variance may contribute additively to phenotypic variance.
- Genetic variance and environmental variance may exhibit interactions and associations (see Figure 26.6, Table 26.3).
- Broad-sense heritability refers to all genetic factors affecting heritability, which includes the additive effects of alleles, effects

due to dominant/recessive relationships, and effects due to epistatic interactions.

- Narrow-sense heritability is heritability that is due to the additive effects of alleles.
- Holt determined that dermal ridge count has a very high heritability value in humans (see Figures 26.7, 26.8).
- Heritability values refer only to particular groups raised in a particular environment (see Table 26.4).

26.4 COMPREHENSION QUESTIONS

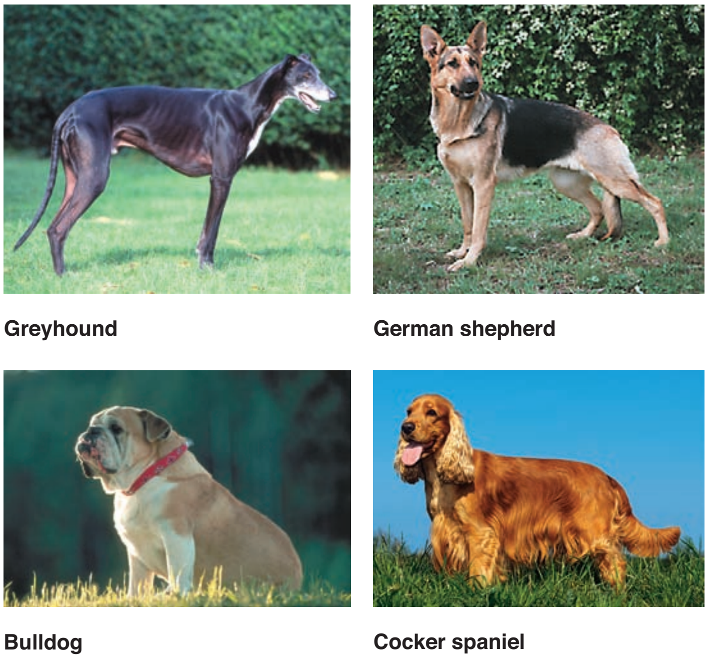
1. In a population of squirrels in North Carolina, the heritability for body weight is high. This means that
   a. body weight is primarily controlled by genes.
   b. the environment has little influence on body weight.
   c. the variation in body weight is mostly due to genetic variation.
   d. both a and b are correct.
2. If two or more different genotypes do not respond to environmental variation in the same way, this is an example of
   a. a genotype-environment association.
   b. a genotype-environment interaction.
   c. the additive effects of alleles.
   d. both a and b.
3. One way to estimate narrow-sense heritability for a given trait is to compare ___________ for __________.
   a. variances, related pairs of individuals
   b. correlation coefficients, related pairs of individuals
   c. variances, unrelated pairs of individuals
   d. correlation coefficients, unrelated pairs of individuals

26.5 SELECTIVE BREEDING

The term **selective breeding** refers to programs and procedures designed to modify phenotypes in species of economically important plants and animals. This phenomenon, also called **artificial selection,** is related to natural selection, discussed in Chapter 25. In forming his theory of natural selection, Charles Darwin was influenced by his observations of selective breeding by pigeon fanciers and other breeders. The primary difference between artificial and natural selection is how the parents are chosen. Natural selection is due to natural variation in reproductive success. In artificial selection, the breeder chooses individuals that possess traits that are desirable from a human perspective. In this section, we will examine the effects of selective breeding and consider its relationship to heritability.

Selective Breeding of Species Can Alter Quantitative Traits Dramatically

For centuries, humans have been practicing selective breeding to obtain domestic species with interesting or agriculturally useful characteristics. For example, the common breeds of dogs have been obtained by selective breeding strategies (**Figure 26.9**). As shown here, it is very striking how selective breeding can modify the quantitative traits in a species. When comparing a greyhound with a bulldog, the magnitude of the differences is fairly amazing. They hardly look like members of the same species. A QTL study in 2007 by Nathan Sutter and colleagues indicates that the size of dogs is determined, in part, by alleles in the *Igf1* gene that encodes a growth hormone called insulin-like growth factor 1. A particular allele of this gene was found to be common to all small breeds of dogs and nearly absent from very large breeds, suggesting that this allele is one of several genes that influences body size in small breeds of dogs.



Greyhound **German shepherd**

Bulldog **Cocker spaniel**

FIGURE 26.9 Some common breeds of dogs that have been obtained by selective breeding.

Genes → Traits By selecting parents carrying the alleles that influence certain quantitative traits in a desired way, dog breeders have produced breeds with distinctive sets of traits. For example, the bulldog has alleles that produce short legs and a flat face. By comparison, the corresponding genes in a German shepherd are found in alleles that produce longer legs and a more pointy snout. All of the dogs shown in this figure carry the same kinds of genes (e.g., many genes that affect their sizes, shapes, and fur color). However, the alleles for many of these genes are different among these dogs, thereby producing breeds with strikingly different phenotypes.

Concept check: What are the similarities and differences between natural selection and selective breeding?

Likewise, most of the food we eat is obtained from species that have been modified profoundly by selective breeding strategies. This includes products such as grains, fruits, vegetables, meat, milk, and juices. **Figure 26.10** illustrates how certain characteristics in the wild mustard plant (*Brassica oleracea*) have been modified by selective breeding to create several varieties of important domesticated crops. This plant is native to Europe and Asia, and plant breeders began to modify its traits approximately 4000 years ago. As seen here, certain quantitative traits in the domestic strains, such as stems and lateral buds, differ considerably from those of the original wild species.

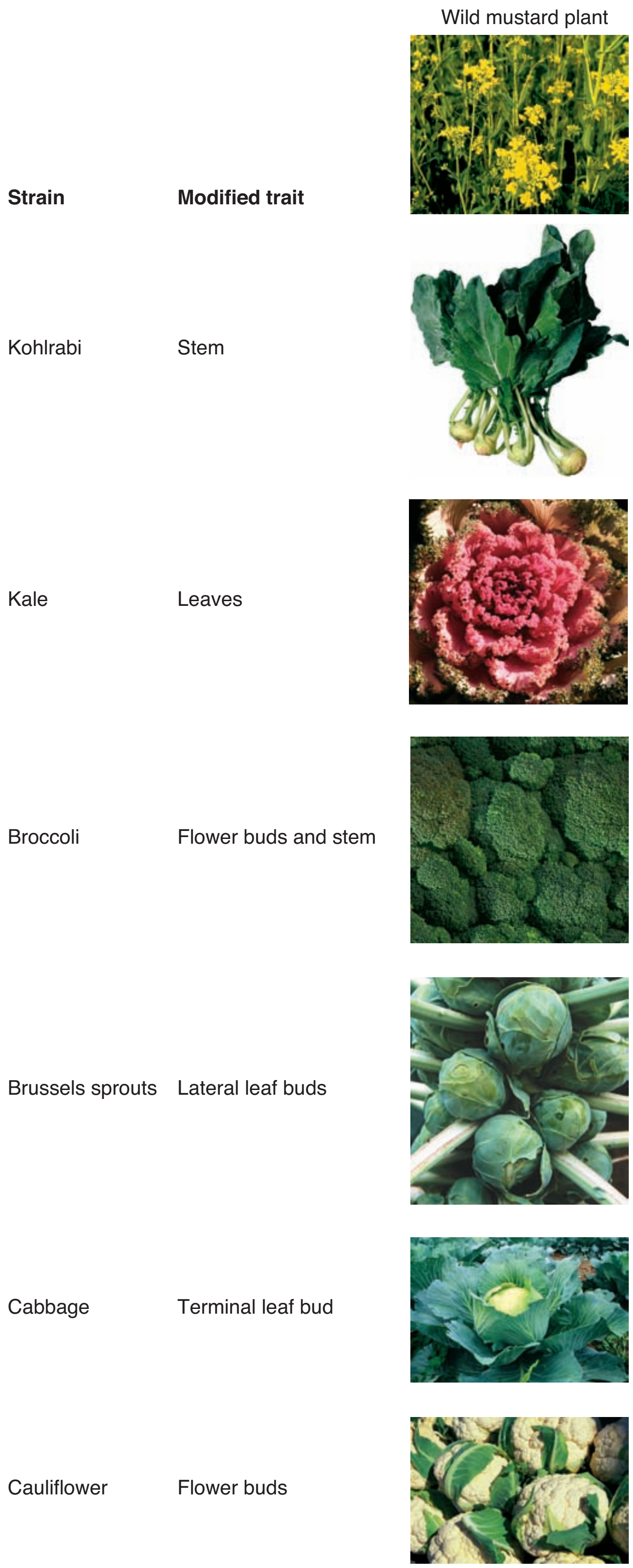


FIGURE 26.10 **Crop plants developed by selective breeding of the wild mustard plant (*Brassica oleracea*).**

Genes → Traits The wild mustard plant carries a large amount of genetic (i.e., allelic) variation, which was used by plant breeders to produce modern strains that are agriculturally desirable and economically important. For example, by selecting for alleles that promote the formation of large lateral leaf buds, the strain of Brussels sprouts was created. By selecting for alleles that alter the leaf morphology, kale was developed. Although these six agricultural plants look quite different from each other, they carry many of the same alleles as the wild mustard. However, they differ in alleles affecting the formation of stems, leaves, flower buds, and leaf buds.

Concept check: Discuss the types of traits that have been subjected to selective breeding in this example.

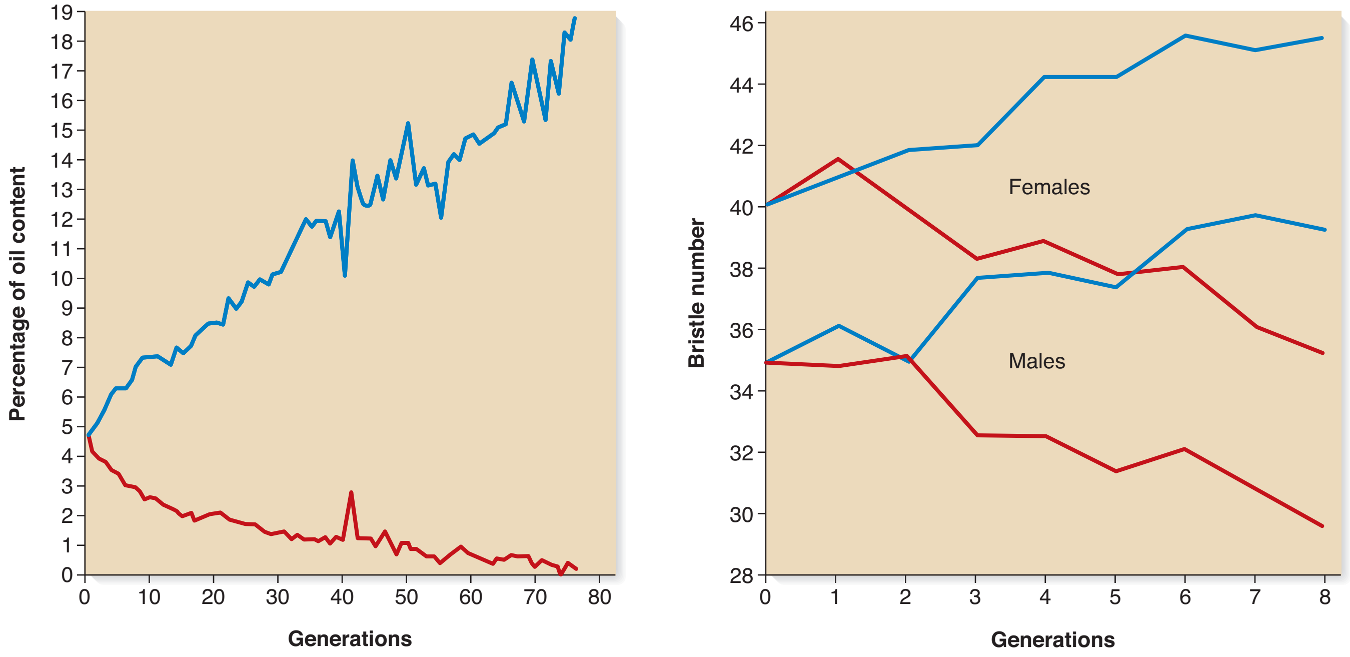
The phenomenon that underlies selective breeding is variation. Within a group of individuals, allelic variation may affect the outcome of quantitative traits. The fundamental strategy of the selective breeder is to choose parents that will pass on alleles to their offspring that produce desirable phenotypic characteristics. For example, if a breeder wants large cattle, the largest members of the herd are chosen as parents for the next generation. These large cattle will transmit an array of alleles to their offspring that confer large size. The breeder often chooses genetically related individuals (e.g., brothers and sisters) as the parental stock. As mentioned previously, the practice of mating between genetically related individuals is known as inbreeding. Some of the consequences of inbreeding are also described in Chapter 25.

What is the outcome when selective breeding is conducted for a quantitative trait? **Figure 26.11a** shows the results of a program begun at the Illinois Agricultural Experiment Station in 1896, even before the rediscovery of Mendel's laws. This experiment began with 163 ears of corn with an oil content ranging from 4% to 6%. In each of 80 succeeding generations, corn plants were divided into two separate groups. In one group, members with the highest oil content were chosen as parents of the next generation. In the other group, members with the lowest oil content were chosen. After 80 generations, the oil content in the first group rose to over 18%; in the other group, it dropped to less than 1%. These results show that selective breeding can modify quantitative traits in a very directed manner.

Similar results have been obtained for many other quantitative traits. **Figure 26.11b** shows an experiment by Kenneth Mather conducted in the 1940s, in which flies were selected on the basis of their bristle number. The starting group had an average of 40 bristles for females and 35 bristles for males. After eight generations, the group selected for high bristle number had an average of 46 bristles for females and 40 for males, whereas the group selected for low bristle number had an average of 36 bristles for females and 30 for males.

When comparing the curves in Figure 26.11, keep in mind that quantitative traits are often at an intermediate value in unselected populations. Therefore, artificial selection can increase or decrease the magnitude of the trait. Oil content can go up or down, and bristle number can increase or decrease. Artificial selection tends to be the most rapid and effective in changing the frequency of alleles that are at intermediate range in a starting population, such as 0.2 to 0.8.

Figure 26.11 also shows the phenomenon known as a **selection limit**—after several generations a plateau is reached where



(a) Results of selective breeding for high and low oil content in corn

(b) Results of selective breeding for high and low bristle number in flies

FIGURE 26.11 **Common results of selective breeding for a quantitative trait.**

Concept check: *What are two reasons why a selection limit is reached in which artificial selection no longer has an effect?*

artificial selection is no longer effective. A selection limit may occur for two reasons. Presumably, the starting population possesses a large amount of genetic variation, which contributes to the diversity in phenotypes. By carefully choosing the parents, each succeeding generation has a higher proportion of the desirable alleles. However, after many generations, the population may be nearly monomorphic for all or most of the desirable alleles that affect the trait of interest. At this point, additional selective breeding will have no effect. When this occurs, the heritability for the trait is near zero, because nearly all genetic variation for the trait of interest has been eliminated from the population. Without the introduction of new mutations into the population, further selection is not possible. A second reason for a selection limit is related to fitness. Some alleles that accumulate in a population due to artificial selection may have a negative influence on the population's overall fitness. A selection limit is reached in which the desired effects of artificial selection are balanced by the negative effects on fitness.

Selective Breeding Provides a Way of Estimating Heritability

Using artificial selection experiments, the response to selection is a common way to estimate the narrow-sense heritability in a starting population. The narrow-sense heritability measured in this way is also called the **realized heritability.** It is calculated as

$$h_N^2 = \frac{R}{S}$$

where

R is the response in the offspring to selection, or the difference between the mean of the offspring and the mean of the starting population.

S is the selection differential in the parents, or the difference between the mean of the parents and the mean of the starting population.

Here,

$$R = \overline{X}_O - \overline{X}$$

$$S = \overline{X}_P - \overline{X}$$

where

$\overline{X}$ is the mean of the starting population

$\overline{X}_O$ is the mean of the offspring

$\overline{X}_P$ is the mean of the parents

So,

$$h_N^2 = \frac{\overline{X}_O - \overline{X}}{\overline{X}_P - \overline{X}}$$

The narrow-sense heritability is the proportion of the variance in phenotype that can be used to predict changes in the population mean when selection is practiced.

As an example, let's suppose we began with a population of fruit flies in which the average bristle number for both sexes was 37.5. The parents chosen from this population had an average bristle number of 40. The offspring of the next generation had an average bristle number of 38.7. With these values, the realized heritability is

$$h_N^2 = \frac{38.7 - 37.5}{40 - 37.5}$$

$$h_N^2 = \frac{1.2}{2.5}$$

$$h_N^2 = 0.48$$

This result tells us that about 48% of the phenotypic variation is due to the additive effects of alleles.

An important aspect of narrow-sense heritabilities is their ability to predict the outcome of selective breeding. In this case, the goal is to predict the mean phenotypes of offspring. If we rearrange our realized heritability equation

$$R = h_N^2 S$$

$$\overline{X}_O - \overline{X} = h_N^2 (\overline{X}_P - \overline{X})$$

This equation is referred to the breeder's equation, because it is used to calculate the mean phenotypes of offspring based on the mean weights of the parents, the mean weights of the starting population, and the heritability. Solved Problem S1 at the end of the chapter illustrates the use of this equation.

26.5 REVIEWING THE KEY CONCEPTS

- Selective breeding refers to programs and procedures designed to modify phenotypes in commercially important plants and animals (see Figures 26.9, 26.10).
- When starting with a genetically diverse population, selective breeding can usually modify a trait in a desired direction until a selection limit is reached (see Figure 26.11).

26.5 COMPREHENSION QUESTIONS

1. For selective breeding to be successful, the starting population must
   a. have genetic variation that affects the trait of interest.
   b. be very large.
   c. be amenable to phenotypic variation caused by environmental effects.
   d. have very little phenotypic variation.
2. The mean weight of cows in a population is 520 kg. Animals with a mean weight of 540 kg are used as parents and produce offspring that have a mean weight of 535 kg. What is the narrow-sense heritability (h_N^2) for body weight in this population of cows?
   a. 0.25
   b. 0.5
   c. 0.75
   d. 1.0

KEY TERMS

Page 646. complex traits, quantitative traits, quantitative genetics
Page 647. continuous traits, meristic traits, threshold traits, discontinuous traits, frequency distribution, normal distribution
Page 648. biometric field, mean
Page 649. variance, standard deviation (*SD*)
Page 650. covariance, correlation coefficient (*r*)
Page 651. regression analysis
Page 652. polygenic inheritance, quantitative trait locus (QTL)
Page 653. QTL mapping
Page 656. heritability
Page 657. inbreeding, monomorphic, genotype-environment interaction
Page 658. genotype-environment association, broad-sense heritability
Page 659. narrow-sense heritability
Page 663. selective breeding, artificial selection
Page 664. selection limit
Page 665. realized heritability
Page 668. regression coefficient

CHAPTER SUMMARY

- Quantitative genetics is the field of genetics concerned with complex and quantitative traits.

26.1 Quantitative Traits

- Quantitative traits can be categorized as anatomical, physiological, or behavioral (see Table 26.1).
- Quantitative traits often exhibit a continuum and may follow a normal distribution (see Figure 26.1).

26.2 Statistical Methods to Evaluate Quantitative Traits

- Statistical methods are used to analyze quantitative traits. These include the mean, variance, standard deviation, covariance, and correlation (see Figure 26.2, Table 26.2).

26.3 Polygenic Inheritance

- Polygenic inheritance refers to an inheritance pattern in which multiple genes affect a single trait.
- The locations along a chromosome that contain genes affecting a quantitative trait are called quantitative trait loci (QTLs).
- Polygenic inheritance and environmental factors may cause a quantitative trait to fall along a continuum (see Figures 26.3, 26.4).
- QTLs are mapped by their linkage to molecular markers (see Figure 26.5).

26.4 Heritability

- Heritability is the amount of phenotypic variation within a group of individuals that is due to genetic variation.

- Genetic variance and environmental variance may contribute additively to phenotypic variance.
- Genetic variance and environmental variance may exhibit interactions and associations (see Figure 26.6, Table 26.3).
- Broad-sense heritability refers to all genetic factors affecting heritability, which includes the additive effects of alleles, effects due to dominant/recessive relationships, and effects due to epistatic interactions.
- Narrow-sense heritability is heritability that is due to the additive effects of alleles.
- Holt determined that dermal ridge count has a very high heritability value in humans (See Figures 26.7, 26.8).
- Heritability values refer only to particular groups raised in a particular environment (see Table 26.4).

26.5 Selective Breeding

- Selective breeding refers to programs and procedures designed to modify phenotypes in commercially important plants and animals (see Figures 26.9, 26.10).
- When starting with a genetically diverse population, selective breeding can usually modify a trait in a desired direction until a selection limit is reached (see Figure 26.11).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. The narrow-sense heritability (h_N^2) for potato weight in a starting population of potatoes is 0.42, and the mean weight is 1.4 pounds. If a breeder crosses plants with average potato weights of 1.9 and 2.1 pounds, respectively, what is the predicted average weight of potatoes in the offspring?

Answer: The mean weight of the parental plants is 2.0 pounds. To solve for the mean weight of the offspring:

$$R = h_N^2 S$$

$$\overline{X}_O - \overline{X} = h_N^2 (\overline{X}_P - \overline{X})$$

$$\overline{X}_O - 1.4 = 0.42\,(2.0 - 1.4)$$

$$\overline{X}_O = 1.65 \text{ pounds}$$

S2. A farmer wants to increase the average body weight in a herd of cattle. She begins with a herd having a mean weight of 595 kg and chooses individuals to breed that have a mean weight of 625 kg. Twenty offspring were obtained, having the following weights in kilograms: 612, 587, 604, 589, 615, 641, 575, 611, 610, 598, 589, 620, 617, 577, 609, 633, 588, 599, 601, and 611. Calculate the realized heritability for body weight in this herd.

Answer:

$$h_N^2 = \frac{R}{S}$$

$$= \frac{\overline{X}_O - \overline{X}}{\overline{X}_P - \overline{X}}$$

We already know the mean weight of the starting herd (595 kg) and the mean weight of the parents (625 kg). The only value missing is the mean weight of the offspring, X_O.

$$\overline{X}_O = \frac{\text{Sum of the offspring's weights}}{\text{Number of offspring}}$$

$$\overline{X}_O = 604 \text{ kg}$$

$$h_N^2 = \frac{604 - 595}{625 - 595}$$

$$= 0.3$$

S3. The following data describe the 6-week weights (in grams) of mice and their offspring of the same sex:

Parent (g)	Offspring (g)
24	26
21	24
24	22
27	25
23	21
25	26
22	24
25	24
22	24
27	24

Calculate the correlation coefficient.

Answer: To calculate the correlation coefficient, we first need to calculate the means and standard deviations for each group:

$$\overline{X}_{parents} = \frac{24 + 21 + 24 + 27 + 23 + 25 + 22 + 25 + 22 + 27}{10} = 24$$

$$\overline{X}_{offspring} = \frac{26 + 24 + 22 + 25 + 21 + 26 + 24 + 24 + 24 + 24}{10} = 24$$

$$SD_{parents} = \sqrt{\frac{0 + 9 + 0 + 9 + 1 + 1 + 4 + 1 + 4 + 9}{9}} = 2.1$$

$$SD_{offspring} = \sqrt{\frac{4 + 0 + 4 + 1 + 9 + 4 + 0 + 0 + 0 + 0}{9}} = 1.6$$

Next, we need to calculate the covariance.

$$C_OV_{(parents,\ offspring)} = \frac{\Sigma\left[(X_P - \overline{X}_P)(X_O - \overline{X}_O)\right]}{N - 1}$$

$$= \frac{0 + 0 + 0 + 3 + 3 + 2 + 0 + 0 + 0 + 0}{9}$$

$$= 0.9$$

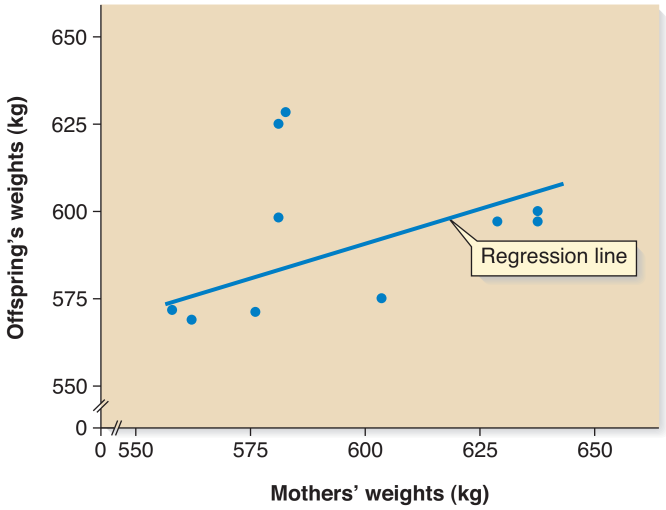
Finally, we calculate the correlation coefficient:

$$r_{(parents,\ offspring)} = \frac{C_OV_{(P,\ O)}}{SD_P\, SD_O}$$

$$r_{(parents,\ offspring)} = \frac{0.9}{(2.1)(1.6)}$$

$$r_{(parent,\ offspring)} = 0.27$$

S4. As described in this chapter, the correlation coefficient provides a way to determine the strength of association between two variables. When the variables are related due to cause and effect (i.e., one variable affects the outcome of another), researchers may use a regression analysis to predict how much one variable changes in response to the other. This is easier to understand if we plot the data for two variables. The graph shown here compares mothers' and offspring's body weights in cattle.



The line running through the data points is called a regression line. It is a line that minimizes the squared vertical distances to all of the points. For many types of data, particularly those involving quantitative traits, the regression line can be represented by the equation

$$Y = bX + a$$

where

b is the regression coefficient

a is a constant

In this example, X is the value of a mother's weight, and Y is the value of its offspring's weight. The value of b, known as the **regression coefficient,** represents the slope of the regression line. The value of a is the y-intercept (i.e., the value of Y when X equals zero). The equation shown is very useful because it allows us to predict the value of Y at any given value of X, and vice versa. To do so, we first need to determine the values of b and a. This can be accomplished in the following manner:

$$b = \frac{C_OV_{(X,Y)}}{V_X}$$

$$a = \overline{Y} - b\overline{X}$$

Once the values of a and b have been computed, we can use them to predict the values of Y or X by using the equation

$$Y = bX + a$$

For example, if $b = 0.5$, $a = 2$, and $X = 58$, this equation can be used to compute a value of Y that equals 31. It is important to keep in mind that this equation predicts the average value of Y. As we see in the preceding figure, the data points tend to be scattered around the regression line. Deviations may occur between the data points and the line. The equation predicts the values that are the most likely to occur. In an actual experiment, however, some deviation will occur between the predicted values and the experimental values due to random sampling error.

Now here is the question. Using the data found in this chapter regarding weight in cattle, what is the predicted weight of an offspring if its mother weighed 660 pounds?

Answer: We first need to calculate a and b.

$$b = \frac{C_OV_{(X,Y)}}{V_X}$$

We need to use the data on page 650 to calculate V_X, which is the variance for the mothers' weights. The variance equals 445.1. The covariance is already calculated on page 650; it equals 152.6.

$$b = \frac{152.6}{445.1}$$

$$b = 0.34$$

$$a = \overline{Y} - b\overline{X}$$

$$a = 598 - (0.34)(596) = 395.4$$

Now we are ready to calculate the predicted weight of the offspring using the equation

$$Y = bX + a$$

In this problem, $X = 660$ pounds

$$Y = 0.34(660) + 395.4$$

$$Y = 619.8 \text{ pounds}$$

The average weight of the offspring is predicted to be 619.8 pounds.

S5. Are the following statements regarding heritability true or false?

A. Heritability applies to a specific population raised in a particular environment.

B. Heritability in the narrow sense takes into account all types of genetic variance.

C. Heritability is a measure of the amount that genetics contributes to the outcome of a trait.

Answer:

A. True

B. False. Narrow-sense heritability considers only the effects of additive alleles.

C. False. Heritability is a measure of the amount of phenotypic variation that is due to genetic variation; it applies to the variation of a specific population raised in a particular environment.

Conceptual Questions

C1. Give several examples of quantitative traits. How are these quantitative traits described within groups of individuals?

C2. At the molecular level, explain why quantitative traits often exhibit a continuum of phenotypes within a population. How does the environment help produce this continuum?

C3. What is a normal distribution? Discuss this curve with regard to quantitative traits within a population. What is the relationship between the standard deviation and the normal distribution?

C4. Explain the difference between a continuous trait and a discontinuous trait. Give two examples of each. Are quantitative traits likely to be continuous or discontinuous? Explain why.

C5. What is a frequency distribution? Explain how the graph is made for a quantitative trait that is continuous.

C6. The variance for weight in a particular herd of cattle is 484 pounds^2. The mean weight is 562 pounds. How heavy would an animal have to be if it was in the top 2.5% of the herd? The bottom 0.13%?

C7. Two different varieties of potatoes both have the same mean weight of 1.5 pounds. One group has a very low variance, and the other has a much higher variance.

A. Discuss the possible reasons for the differences in variance.

B. If you were a potato farmer, would you rather raise a variety with a low or high variance? Explain your answer from a practical point of view.

C. If you were a potato breeder and you wanted to develop potatoes with a heavier weight, would you choose the variety with a low or high variance? Explain your answer.

C8. If an r value equals 0.5 and $N = 4$, would you conclude a positive correlation is found between the two variables? Explain your answer. What if $N = 500$?

C9. What does it mean when a correlation coefficient is negative? Can you think of examples?

C10. When a correlation coefficient is statistically significant, what do you conclude about the two variables? What do the results mean with regard to cause and effect?

C11. What is polygenic inheritance? Discuss the issues that make polygenic inheritance difficult to study.

C12. What is a quantitative trait locus (QTL)? Does a QTL contain one gene or multiple genes? What technique is commonly used to identify QTLs?

C13. Let's suppose that weight in a species of mammal is polygenic, and each gene exists as a heavy and light allele. If the allele frequencies in the population were equal for both types of allele (i.e., 50% heavy alleles and 50% light alleles), what percentage of individuals would be homozygous for the light alleles at all of the genes affecting this trait, if the trait was determined by the following number of genes?

A. Two

B. Three

C. Four

C14. The broad-sense heritability for a trait equals 1.0. In your own words, explain what this value means. Would you conclude that the environment is unimportant in the outcome of this trait? Explain your answer.

C15. From an agricultural point of view, discuss the advantages and disadvantages of selective breeding. It is common for plant breeders to take two different, highly inbred strains, which are the product of many generations of selective breeding, and cross them to make hybrids. How does this approach overcome some of the disadvantages of selective breeding?

C16. Many beautiful varieties of roses have been produced, particularly in the last few decades. These newer varieties often have very striking and showy flowers, making them desirable as horticultural specimens. However, breeders and novices alike have noticed that some of these newer varieties do not have very fragrant flowers compared with the older, more traditional varieties. From a genetic point of view, suggest an explanation why some of these newer varieties with superb flowers are not as fragrant.

C17. In your own words, explain the meaning of the term *heritability.* Why is a heritability value valid only for a particular population of individuals raised in a particular environment?

C18. What is the difference between broad-sense heritability and narrow-sense heritability? Why is narrow-sense heritability such a useful concept in the field of agricultural genetics?

C19. The heritability for egg weight in a group of chickens on a farm in Maine is 0.95. Are the following statements regarding heritability true or false? If a statement is false, explain why.

A. The environment in Maine has very little effect on the outcome of this trait.

B. Nearly all of the phenotypic variation for this trait in this group of chickens is due to genetic variation.

C. The trait is polygenic and likely to involve a large number of genes.

D. Based on the observation of the heritability in the Maine chickens, it is reasonable to conclude that the heritability for egg weight in a group of chickens on a farm in Montana is also very high.

C20. In a fairly large population of people living in a commune in the southern United States, everyone cares about good nutrition. All of the members of this population eat very nutritious foods, and their diets are very similar to each other. How do you think the height of individuals in this commune population would compare with that of the general population in the following categories?

A. Mean height

B. Heritability for height

C. Genetic variation for alleles that affect height

C21. When artificial selection is practiced over many generations, it is common for the trait to reach a plateau in which further selection has little effect on the outcome of the trait. This phenomenon is illustrated in Figure 26.11. Explain why.

C22. Discuss whether a natural population of wolves or a domesticated population of German shepherds is more likely to have a higher heritability for the trait of size.

Application and Experimental Questions

E1. Here are data for height and weight among 10 male college students.

Height (cm)	Weight (kg)
159	48
162	50
161	52
175	60
174	64
198	81
172	58
180	74
161	50
173	54

A. Calculate the correlation coefficients for this group.

B. Is the correlation coefficient statistically significant? Explain.

E2. The abdomen length (in millimeters) was measured in 15 male *Drosophila,* and the following data were obtained: 1.9, 2.4, 2.1, 2.0, 2.2, 2.4, 1.7, 1.8, 2.0, 2.0, 2.3, 2.1, 1.6, 2.3, and 2.2. Calculate the mean, standard deviation, and variance for this population of male fruit flies.

E3. In one strain of cabbage, you conduct an RFLP analysis of head weight; you determine that seven QTLs affect this trait. In another strain of cabbage, you find that only four QTLs affect this trait. Note that both strains of cabbage are from the same species, although they may have been subjected to different degrees of inbreeding. Explain how one strain can have seven QTLs and another strain four QTLs for exactly the same trait. Is the second strain missing three genes?

E4. From an experimental viewpoint, what does it mean to say that an RFLP is associated with a trait? Let's suppose that two strains of pea plants differ in two RFLPs that are linked to two genes governing pea size. RFLP-1 is found in 2000-bp and 2700-bp bands, and RFLP-2 is found in 3000-bp and 4000-bp bands. The plants producing large peas have RFLP-1 (2000 bp) and RFLP-2 (3000 bp); those producing small peas have RFLP-1 (2700 bp) and RFLP-2 (4000 bp). A cross is made between these two strains, and the F_1 offspring are allowed to self-fertilize. Five phenotypic classes are observed: small peas, small-medium peas, medium peas, medium-large peas, and large peas. We assume that each of the two genes makes an equal contribution to pea size and that the genetic variance is additive. Draw a gel and explain what RFLP banding patterns you would expect to observe for these five phenotypic categories. Note: Certain phenotypic categories may have more than one possible banding pattern.

E5. Let's suppose that two strains of pigs differ in 500 RFLPs. One strain is much larger than the other. The pigs are crossed to each other, and the members of the F_1 generation are also crossed among themselves to produce an F_2 generation. Three distinct RFLPs are associated with F_2 pigs that are larger. How would you interpret these results?

E6. Outline the steps you would follow to determine the number of genes that influence the yield of rice. Describe the results you might get if rice yield is governed by variation in six different genes.

E7. In a wild strain of tomato plants, the phenotypic variance for tomato weight is 3.2 g^2. In another strain of highly inbred tomatoes raised under the same environmental conditions, the phenotypic variance is 2.2 g^2. With regard to the wild strain,

A. Estimate V_G.

B. What is $h_B{}^2$?

C. Assuming that all of the genetic variance is additive, what is $h_N{}^2$?

E8. The average thorax length in a *Drosophila* population is 1.01 mm. You want to practice selective breeding to make larger *Drosophila*. To do so, you choose 10 parents (5 males and 5 females) of the following sizes: 0.97, 0.99, 1.05, 1.06, 1.03, 1.21, 1.22, 1.17, 1.19, and 1.20. You mate them and then analyze the thorax sizes of 30 offspring (half male and half female):

0.99, 1.15, 1.20, 1.33, 1.07, 1.11, 1.21, 0.94, 1.07, 1.11, 1.20, 1.01, 1.02, 1.05, 1.21, 1.22, 1.03, 0.99, 1.20, 1.10, 0.91, 0.94, 1.13, 1.14, 1.20, 0.89, 1.10, 1.04, 1.01, 1.26

Calculate the realized heritability in this group of flies.

E9. In a strain of mice, the average 6-week body weight is 25 g and the narrow-sense heritability for this trait is 0.21.

A. What would be the average weight of the offspring if parents with a mean weight of 27 g were chosen?

B. What weight of parents would you have to choose to obtain offspring with an average weight of 26.5 g?

E10. You need to understand Solved Problem S4 before answering this question. The variance in height for fathers (in square inches) was 112, the variance for sons was 122, and the covariance was 144. The mean height for fathers was 68 inches, and the mean height for sons was 69 inches. If a father had a height of 70 inches, what is the most probable height of his son?

E11. A danger in computing heritability values from studies involving genetically related individuals is the possibility that these individuals share more similar environments than do unrelated individuals. In the experiment of Figure 26.8, which data are the most compelling evidence that ridge count is not caused by genetically related individuals sharing common environments? Explain.

E12. A large, genetically heterogeneous group of tomato plants was used as the original breeding stock by two different breeders, named Mary and Hector. Each breeder was given 50 seeds and began an artificial selection strategy, much like the one described in Figure 26.11. The seeds were planted, and the breeders selected the 10 plants with the highest mean tomato weights as the breeding stock for the next generation. This process was repeated over the course of 12 growing seasons, and the following data were obtained:

	Mean Weight of Tomatoes (pounds)	
Year	Mary's Tomatoes	Hector's Tomatoes
1	0.7	0.8
2	0.9	0.9
3	1.1	1.2
4	1.2	1.3
5	1.3	1.3
6	1.4	1.4
7	1.4	1.5
8	1.5	1.5
9	1.5	1.5
10	1.5	1.5
11	1.5	1.5
12	1.5	1.5

A. Explain these results.

B. Another tomato breeder, named Martin, got some seeds from Mary's and Hector's tomato strains (after 12 generations), grew the plants, and then crossed them to each other. The mean weight of the tomatoes in these hybrids was about 1.7 pounds. For a period of 5 years, Martin subjected these hybrids to the same experimental strategy that Mary and Hector had followed, and he obtained the following results:

Mean Weight of Tomatoes (pounds)

Year	Martin's Tomatoes
1	1.7
2	1.8
3	1.9
4	2.0
5	2.0

Explain Martin's data. Why was Martin able to obtain tomatoes heavier than 1.5 pounds, whereas Mary's and Hector's strains appeared to plateau at this weight?

E13. The correlations for height were determined for 15 pairs of individuals with the following genetic relationships:

Mother/daughter: 0.36

Mother/granddaughter: 0.17

Sister/sister: 0.39

Sister/sister (fraternal twins): 0.40

Sister/sister (identical twins): 0.77

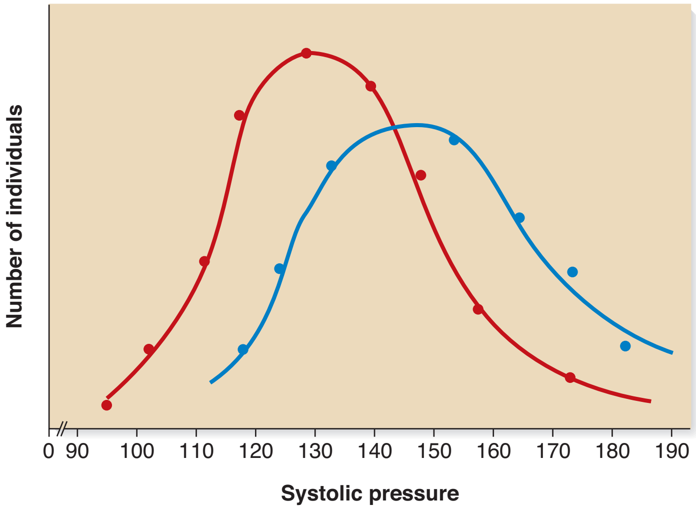
What is the average heritability for height in this group of females?

E14. An animal breeder had a herd of sheep with a mean weight of 254 pounds at 3 years of age. He chose animals with mean weights of 281 pounds as parents for the next generation. When these offspring reached 3 years of age, their mean weights were 269 pounds.

A. Calculate the narrow-sense heritability for weight in this herd.

B. Using the heritability value that you calculated in part A, what weight of animals would you have to choose to get offspring that weigh 275 pounds (at 3 years of age)?

E15. The trait of blood pressure in humans has a frequency distribution that is similar to a normal distribution. The following graph shows the ranges of blood pressures for a selected population of people. The red line depicts the frequency distribution of the systolic pressures for the entire population. Several individuals with high blood pressure were identified, and the blood pressures of their relatives were determined. This frequency distribution is depicted with a blue line. (Note: The blue line does not include the people who were identified with high blood pressure; it includes only their relatives.)



What do these data suggest with regard to a genetic basis for high blood pressure? What statistical approach could you use to determine the heritability for this trait?

Questions for Student Discussion/Collaboration

1. Discuss why heritability is an important phenomenon in agriculture. Discuss how it is misunderstood.
2. From a biological viewpoint, speculate as to why many traits seem to fit a normal distribution. Students with a strong background in math and statistics may want to explain how a normal distribution is generated, and what it means. Can you think of biological examples that do not fit a normal distribution?

Answers to Comprehension Questions

26.1: d, c

26.2: d, b

26.3: d, b, c

26.4: c, b, b

26.5: a, c

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

27

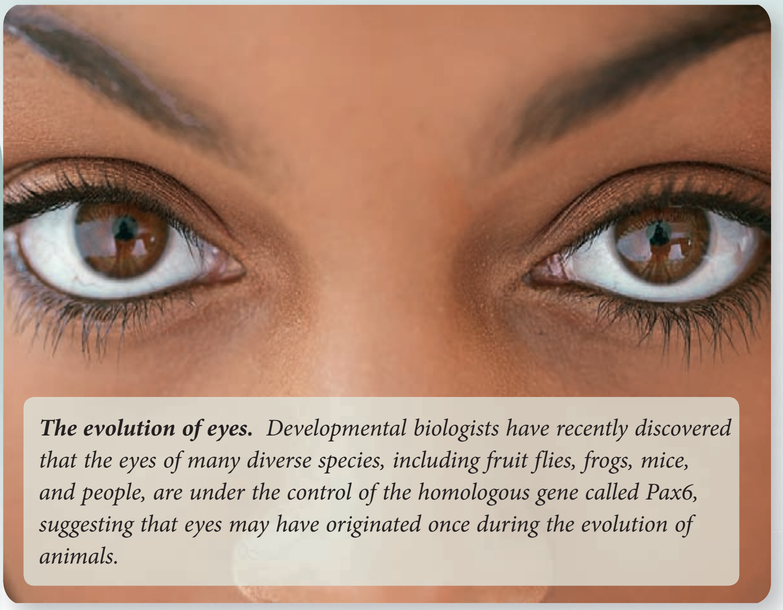
CHAPTER OUTLINE

27.1 Origin of Species

27.2 Phylogenetic Trees

27.3 Molecular Evolution and Molecular Clocks

27.4 Evo-Devo: Evolutionary Developmental Biology



The evolution of eyes. *Developmental biologists have recently discovered that the eyes of many diverse species, including fruit flies, frogs, mice, and people, are under the control of the homologous gene called Pax6, suggesting that eyes may have originated once during the evolution of animals.*

EVOLUTIONARY GENETICS

Biological evolution, or simply **evolution,** is a heritable change in one or more characteristics of a population or species from one generation to the next. Evolution can be viewed on a small scale as it relates to a single gene or it can be viewed on a larger scale as it relates to the formation of new species. In Chapter 25, we examined several factors that cause allele frequencies to change in populations. This process, known as **microevolution,** concerns the changing composition of gene pools with regard to particular alleles over measurable periods of time. As we have seen, several evolutionary mechanisms, such as mutation, natural selection, genetic drift, migration, and nonrandom mating, affect the allele and genotypic frequencies within natural populations. On a microevolutionary scale, evolution can be viewed as a change in allele frequency over time.

The term **macroevolution** refers to large-scale evolutionary changes that lead to the emergence of new species and higher taxa. It concerns the establishment of the diversity of organisms over long periods of time through the accumulated evolution and extinction of many species. The goal of this chapter is to relate phenotypic changes that occur during evolution to the underlying genetic changes that cause them to happen. In the first part of the chapter, we will be concerned with the origin of new species. The question of how species form has been central to the development of evolutionary theory.

In the second part of this chapter, we will link molecular genetics to the evolution of species. The techniques for analyzing chromosomes and DNA sequences have greatly enhanced our understanding of evolutionary processes at the molecular level. The term **molecular evolution** refers to patterns and processes associated with evolutionary change at the molecular level. Such changes may be phenotypically neutral or they may underlie the phenotypic changes associated with evolution. We will explore how molecular data can provide information about the evolutionary relationships among different organisms. Finally, in the last part of the chapter, we will see how genes that affect embryonic development can have a dramatic effect on the phenotypes of organisms. The field of evolutionary developmental biology (evo-devo) focuses on the role of developmental genes in the formation of traits that are important in the evolution of new species.

The topics of molecular evolution and evo-devo are a fitting way to end our discussion of genetics because they integrate the ongoing theme of this textbook—the relationship between molecular genetics and traits—in the broadest and most profound ways. Theodosius Dobzhansky, an influential evolutionary scientist, once said, "Nothing in biology makes sense except in the light of evolution." The extraordinarily diverse and seemingly bizarre array of species on our planet can be explained naturally within the context of evolution. An examination of molecular evolution allows us to make sense of the existence of these species at both the population and the molecular levels.

27.1 ORIGIN OF SPECIES

Charles Darwin, a British naturalist born in 1809, proposed the theory of evolution and provided evidence that existing species have evolved from preexisting ones. Like many great scientists, Darwin had a broad background in science, which enabled him to see connections among different disciplines. His thinking was influenced by the field of geology. According to Charles Lyell, the processes that alter the Earth are uniform through time. This view, which was known as uniformitarianism, suggested that the Earth is very old and that slow geological processes can lead eventually to substantial changes in the Earth's characteristics.

Darwin's own experimental observations also greatly affected his thinking. His famous voyage on the *HMS Beagle*, which lasted from 1832 to 1836, involved a careful examination of many different species. He observed the similarities among many discrete species, yet noted the differences that enabled them to be adapted to their environmental conditions. He was particularly struck by the distinctive adaptations of island species. For example, the finches found on the Galápagos Islands had unique phenotypic characteristics compared with those of similar finches found on the mainland.

A third important influence on Darwin was the book *An Essay on the Principle of Population*, published in 1798 by Thomas Malthus, an English economist. Malthus asserted that the size of human populations can, at best, increase arithmetically due to increased land usage and improvements in agriculture, whereas the reproductive potential of humans can increase geometrically. He argued that famine, war, and disease limit population growth, especially among the poor.

With these three ideas in mind, Darwin had largely formulated his theory of evolution by natural selection by the mid-1840s. The geologist Charles Lyell, who had greatly influenced Darwin's thinking, strongly encouraged Darwin to publish his theory of evolution. In 1856, Darwin began to write a long book to explain his ideas. In 1858, however, Alfred Wallace, a naturalist working in the East Indies, sent Darwin an unpublished manuscript to read prior to its publication. In it, Wallace proposed the same ideas concerning evolution. Darwin therefore quickly excerpted some of his own writings on this subject, and two papers, one by Darwin and one by Wallace, were published in the *Proceedings of the Linnaean Society of London*. These papers were not widely recognized. In 1859, however, Darwin published his book, *On the Origin of Species by Means of Natural Selection*, which expounded his ideas in greater detail and with experimental support. This book, which received high praise from many scientists and scorn from others, started a great debate concerning natural selection. Although some of his ideas were incomplete because the genetic basis of traits was not understood at that time, Darwin's work represents one of the most important contributions to our understanding of biology.

Darwin called evolution "the theory of descent with modification through variation and natural selection." This form of evolution is sometimes called **adaptive evolution.** It is based on two fundamental principles: genetic variation and natural selection. A modern interpretation of adaptive evolution can view these two principles at the species level (macroevolution) and at the level of genes in populations (microevolution).

1. *Genetic variation at the species level:* As we saw in Chapter 25, genetic variation is a consistent feature of natural populations. Darwin observed that many species exhibit a great amount of phenotypic variation. Although the theory of evolution preceded Mendel's pioneering work in genetics, Darwin (as well as many other people before him) observed that offspring resemble their parents more than they do unrelated individuals. Therefore, he assumed that traits are passed from parent to offspring. However, the genetic basis for the inheritance of traits was not understood at that time.

   At the gene level: Genetic variation can involve allelic differences in genes. These differences are caused by random mutations. Alternative alleles may affect the functions of the proteins they encode, thereby affecting the phenotype of the organism. Likewise, changes in chromosome structure and number may affect gene expression, thereby influencing the phenotype of the individual.
2. *Natural selection at the species level:* Darwin agreed with Malthus that most species produce many more offspring than will survive and reproduce, resulting in an ever-present struggle for existence. Over the course of many generations, those individuals who happen to possess the most favorable traits will dominate the composition of the population. The result of natural selection is to make a species better adapted to its environment and more successful at reproduction.

   At the gene level: Some alleles encode proteins that provide the individual with a selective advantage. Over time, natural selection may change the allele frequencies of genes, thereby leading to the fixation of beneficial alleles and the elimination of detrimental alleles.

In this section, we will consider the different characteristics that biologists examine when deciding if two groups of organisms constitute different species and explore the features of evolution as it occurs in natural populations over time.

Each Species Is Established Using Characteristics and Histories That Distinguish It from Other Species

Before we begin to consider how biologists study the evolution of new species, we need to consider how species are defined and identified. A **species** refers to a group of organisms that maintains a distinctive set of attributes in nature. How many different species are on Earth? The number is astounding. A study done by biologist Edward O. Wilson and colleagues in 1990 estimated the known number of species at approximately 1.4 million. Currently, about 1.75 million species have been identified. However, a vast number of species have yet to be established. This is particularly true among prokaryotic organisms, which are difficult to categorize into distinct species. Common estimates of the total number of species range from 5 to 50 million!

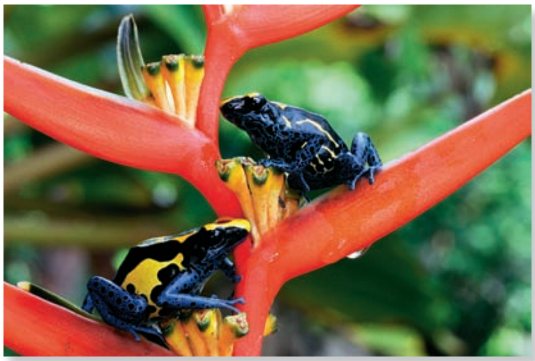
When studying natural populations, evolutionary biologists are often confronted with situations in which some differences between two populations are apparent, but it is difficult to decide whether the two populations truly represent separate species. When two or more geographically restricted groups of the same species display one or more traits that are somewhat different but not enough to warrant their placement into different species, biologists sometimes classify such groups as **subspecies.** Similarly, many bacterial species are subdivided into ecotypes. Each **ecotype** is a genetically distinct population adapted to its local environment.

Members of the same species share an evolutionary history that is distinct from other species. Although this may seem like a reasonable way to characterize a given species, most evolutionary biologists agree that the identification of many species is a difficult undertaking. What criteria do we use to distinguish species? How many differences must exist between two populations to classify them as distinct species? Such questions are often difficult to answer.

The characteristics that a biologist uses to identify a species depend, in large part, on the species in question. For example, the traits used to distinguish insect species are quite different from those used to identify different bacterial species. The relatively high level of horizontal gene transfer among bacteria presents special challenges in the grouping of bacterial species. Among bacteria, it is sometimes very difficult and perhaps arbitrary to divide closely related organisms into separate species. The most commonly used characteristics for identifying species are morphological traits, the ability to interbreed, molecular features, ecological factors, and evolutionary relationships. A comparison of these concepts will help you to appreciate the various approaches that biologists use to identify the bewildering array of species on our planet.

Morphological Traits One way to establish that a population constitutes a unique species is based on their physical characteristics. Organisms are classified as the same species if their anatomical traits appear to be very similar. Likewise, microorganisms can be classified according to morphological traits at the cellular level. By comparing many different morphological traits, biologists may decide that certain populations constitute a unique species.

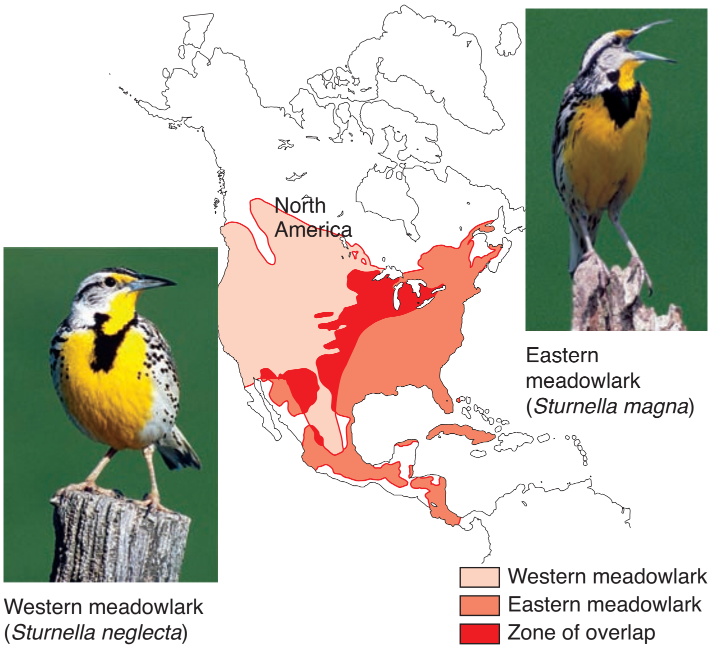
Although an analysis of morphological traits is a common way for biologists to establish that a particular group constitutes a species, it has a few drawbacks. First, it may be difficult to decide how many traits to consider. In addition, it is difficult to analyze quantitative traits, such as size and weight, which vary in a continuous way among members of the same species. Another drawback is that the degree of dissimilarity that distinguishes different species may not show a simple relationship; the members of the same species sometimes look very different, and conversely, members of different species sometimes look remarkably similar to each other. For example, **Figure 27.1a** shows two different frogs of the species *Dendrobates tinctorius*, commonly called the dyeing poison frog. This species exists in many different-colored morphs, which are individuals of the same species that have noticeably dissimilar appearances. In contrast, **Figure 27.1b** shows two different species



(a) Frogs of the same species

FIGURE 27.1 **Morphological dissimilarities and similarities in species.** **(a)** In some cases, two members of the same species can look quite different. This photograph shows two members of the same species, the dyeing poison frog (*Dendrobates tinctorius*). **(b)** In comparison, members of different species can look quite similar as illustrated by the western meadowlark (*Sturnella neglecta*) and eastern meadowlark (*Sturnella magna*).

Concept check: *Does this figure illustrate the strength or potential problem of using morphological traits to establish species?*


North America
Eastern meadowlark (*Sturnella magna*)
Western meadowlark (*Sturnella neglecta*)
Western meadowlark
Eastern meadowlark
Zone of overlap

(b) Birds of different species

of meadowlarks: the western meadowlark (*Sturnella neglecta*) and eastern meadowlark (*Sturnella magna*). Both species are nearly identical in shape, coloration, and habitat, and their ranges overlap in the central United States.

Reproductive Isolation Why do biologists describe two species, such as the western and eastern meadowlarks, as being different if they are morphologically similar? One reason is that biologists have discovered that they are unable to breed with each other in nature. In the zone of overlap, very little interspecies mating takes place between the two species, largely due to differences in their songs. The song of the western meadowlark is a long series of flutelike gurgling notes that go down the scale. By comparison, the eastern meadowlark's song is a simple series of whistles, typically about four or five notes. These differences in songs enable meadowlarks to recognize potential mates as members of their own species.

Therefore, a second way to identify a species is by the ability to interbreed. In the late 1920s, geneticist Theodosius Dobzhansky proposed that each species is reproductively isolated from other species. Such **reproductive isolation** prevents one species from successfully interbreeding with other species. In 1942, evolutionary biologist Ernst Mayr expanded on the ideas of Dobzhansky to provide a definition of a species. According to Mayr, a key feature of sexually reproducing species is that, in nature, the members of one species have the potential to interbreed with one another to produce viable, fertile offspring but cannot successfully interbreed with members of other species.

Reproductive isolation has been used to distinguish many plant and animal species, especially those, like the meadowlarks, that look alike but do not interbreed. How does reproductive isolation occur? Table 27.1 describes several ways. These are classified as **prezygotic isolating mechanisms,** which prevent the formation of a zygote, and **postzygotic isolating mechanisms,** which prevent the development of a viable and fertile individual after fertilization has taken place. Reproductive isolation in nature may be circumvented when species are kept in captivity. For example, different species of the genus *Drosophila* rarely mate with each other in nature. In the laboratory, however, it is fairly easy to produce interspecies hybrids.

Although reproductive isolation has been commonly used to classify species, it suffers from four main problems. First, in nature, it may be difficult to determine if two populations are reproductively isolated, particularly if they are populations with nonoverlapping geographical ranges. Second, biologists have noted many cases in which two different species can interbreed in nature yet consistently maintain themselves as separate species. For example, different species of yucca plants, such as *Yucca pallida* and *Yucca constricta*, do interbreed in nature yet typically maintain populations with distinct characteristics. For this reason, they are viewed as distinct species. A third drawback of reproductive isolation is that it does not apply to asexual species such as bacteria. Likewise, some species of plants and fungi only reproduce asexually. Finally, a fourth drawback is that it cannot be applied to the identification of extinct species. For these reasons, reproductive isolation has been primarily used to distinguish between closely related species of modern animals and plants that reproduce sexually.

TABLE **27.1**

Types of Reproductive Isolation Among Different Species

Prezygotic Isolating Mechanisms	
Habitat isolation	Species may occupy different habitats so they never come in contact with each other.
Temporal isolation	Species have different mating or flowering seasons, mate at different times of day, or become sexually active at different times of the year.
Sexual isolation	Sexual attraction between males and females of different animal species is limited due to differences in behavior, physiology, or morphology.
Mechanical isolation	The anatomical structures of genitalia prevent mating between different species.
Gametic isolation	Gametic transfer takes place, but the gametes fail to unite with each other. This can occur because the male and female gametes fail to attract, because they are unable to fuse, or because the male gametes are inviable in the female reproductive tract of another species.
Postzygotic Isolating Mechanisms	
Hybrid inviability	The egg of one species is fertilized by the sperm from another species, but the fertilized egg fails to develop past early embryonic stages.
Hybrid sterility	The interspecies hybrid survives, but it is sterile. For example, the mule, which is sterile, is a cross between a female horse (*Equus caballus*) and a male donkey (*Equus asinus*).
Hybrid breakdown	The F_1 interspecies hybrid is viable and fertile, but succeeding generations (i.e., F_2, etc.) become increasingly inviable. This is usually due to the formation of less fit genotypes by genetic recombination.

Molecular Features Molecular features are now commonly used to determine if two different populations are different species. Evolutionary biologists often compare DNA sequences within genes, gene order along chromosomes, chromosome structure, and chromosome number as features to identify similarities and differences among different populations. DNA sequence differences are often used to compare populations. For example, researchers may compare the DNA sequence of the 16S rRNA gene between different bacterial populations as a way to decide if the two populations represent different species. When the sequences are very similar, such populations are typically judged to be the same species. However, it may be difficult to draw the line when separating groups into different species. Is a 2% difference in their genome sequences sufficient to warrant placement into two different species, or do we need a 5% difference?

Ecological Factors A variety of factors related to an organism's habitat can be used to distinguish one species from another.

For example, certain species of warblers can be distinguished by the habitat in which they forage for food. Some species search the ground for food, others forage in bushes or small trees, and some species primarily forage in tall trees. Such habitat differences can be used to distinguish different species that look morphologically similar.

Many bacterial species have been categorized as distinct species based on ecological factors. Bacterial cells of the same species are likely to use the same types of resources (e.g., sugars and vitamins) and grow under the same types of conditions (e.g., temperature and pH). However, a drawback of this approach is that different groups of bacteria sometimes display very similar growth characteristics, and even the same species may show great variation in the growth conditions it will tolerate.

Evolutionary Relationships Later in this chapter, we will examine the methods used to produce evolutionary trees that describe the evolutionary relationships among different species. In some cases, such relationships are based on an analysis of the fossil record. Alternatively, another way to establish evolutionary relationships is by the analysis of DNA sequences. Researchers can obtain samples of cells from different individuals and compare the genes within those cells to see how similar or different they are.

Species Concepts Thus far, we have considered multiple ways of identifying species. The most commonly used characteristics for distinguishing species are morphological traits, reproductive isolation, molecular features, ecological factors, and evolutionary relationships.

A **species concept** is a way to define the concept of a species and/or provide an approach for distinguishing one species from another. In 1942, Ernst Mayr proposed an early species concept, called the **biological species concept.** According to this idea, a species is a group of individuals whose members have the potential to interbreed with one another in nature to produce viable, fertile offspring but cannot successfully interbreed with members of other species. The biological species concept emphasizes reproductive isolation as the most important criterion for delimiting species. Since then, over 20 different species concepts have been proposed by a variety of evolutionary biologists. Another example is the **evolutionary lineage concept** proposed by American paleontologist George Gaylord Simpson in 1961. According to this idea, a species should be defined based on the separate evolution of lineages. A third example is the **ecological species concept,** described by American evolutionary biologist Leigh Van Valen in 1976. According to this viewpoint, each species occupies an ecological niche, which is the unique set of habitat resources that a species requires, as well as its influence on the environment and other species.

Most evolutionary biologists agree that different methods are needed to distinguish the vast array of species on Earth. Even so, some evolutionary biologists have questioned whether it is valid to have many different species concepts. In 1998, Kevin de Queiroz suggested that there is only a single general species concept, which concurs with Simpson's evolutionary lineage concept and includes all previous concepts. According to de Queiroz's **general lineage concept,** each species is a population of an independently evolving lineage. Each species has evolved from a specific series of ancestors and, as a consequence, forms a group of organisms with a particular set of characteristics. Multiple criteria are used to determine if a population is part of an independent evolutionary lineage, and thus a species, which is distinct from others. Typically, researchers use analyses of morphology, reproductive isolation, DNA sequences, and ecology to determine if a population or group of populations is distinct from others. Because of its generality, the general lineage concept has received significant support.

Speciation Usually Occurs via a Branching Process Called Cladogenesis

Speciation is the process by which new species are formed via evolution. One way it can occur is via **anagenesis** (from the Greek *ana,* up, and *genesis,* origin), in which a single species evolves into a different species over the course of many generations (**Figure 27.2a**). However, most evolutionary biologists would argue that anagenesis is not a common mechanism for speciation, though it could occur if an entire species was confined to a single location for long periods of time.

By comparison, speciation primarily occurs via **cladogenesis** (from the Greek *clados,* branch), which involves the division of a species into two or more species (**Figure 27.2b**). This form of speciation increases species diversity. Although cladogenesis is usually thought of as a splitting process, it commonly occurs as a budding process, which results in the original species plus one or more new species with different characteristics. If we view evolution as a tree, the new species bud from the original species and develop characteristics that prevent them from breeding with the original one.

Cladogenesis Can Be Allopatric, Parapatric, or Sympatric

Depending on the geographic locations of the evolving population(s) and the environment that a species occupies, cladogenesis is categorized as allopatric, parapatric, or sympatric (**Table 27.2**).

Allopatric Speciation **Allopatric speciation** (from the Greek *allos*, other, and Latin *patria*, homeland) is thought to be the most prevalent way for a species to diverge. It happens when members of a species become geographically separated from the other members. This form of speciation can occur by the geographic subdivision of large populations via geological processes. For example, a mountain range may emerge and split a species that occupies the lowland regions, or a creeping glacier may divide a population. **Figure 27.3** shows an interesting example in which geological separation promoted speciation. Two species of antelope squirrels occupy opposite rims of the Grand Canyon. On the south rim is Harris's antelope squirrel (*Ammospermophilus harrisi*), whereas a closely related white-tailed antelope squirrel (*Ammospermophilus leucurus*) is found on the north rim. Presumably, these two species evolved from a common species that existed before the canyon was formed. Over time, the

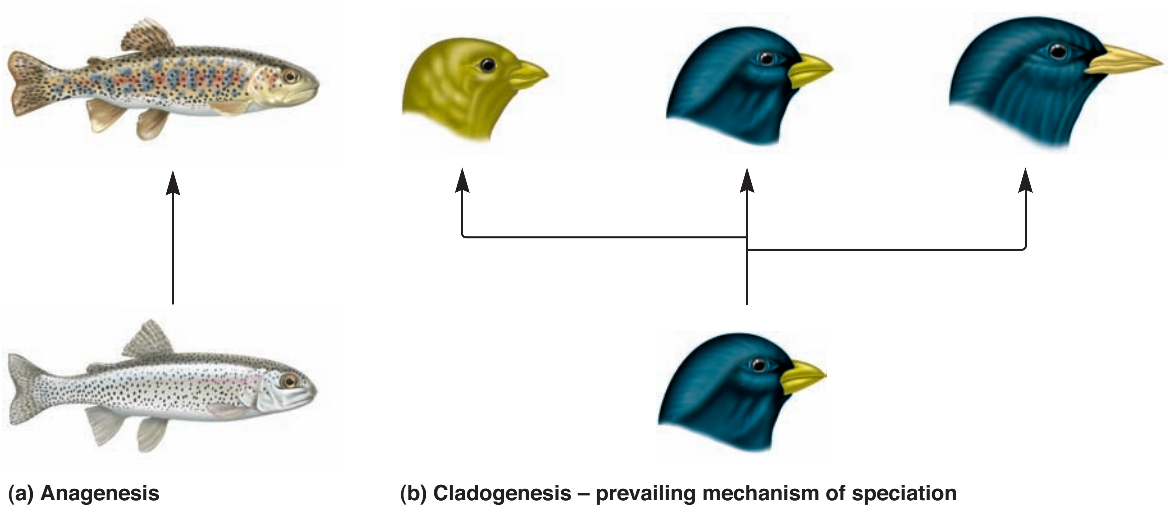

(a) Anagenesis
(b) Cladogenesis – prevailing mechanism of speciation

FIGURE 27.2 **A comparison between anagenesis and cladogenesis: two patterns of speciation.** **(a)** Anagenesis is the change of one species into another. **(b)** Cladogenesis involves a process in which one original species is separated into two or more different species.

Concept check: Which is more common: anagenesis or cladogenesis?

TABLE 27.2

Common Genetic Mechanisms That Underlie Allopatric, Parapatric, and Sympatric Speciation

Type of Speciation	Common Genetic Mechanisms Responsible for Speciation
Allopatric—two large populations are separated by geographic barriers	Many small genetic differences may accumulate over a long period, leading to reproductive isolation. Some of these genetic differences may be adaptive, whereas others are neutral.
Allopatric—a small founding population separates from the main population	Genetic drift may lead to the rapid formation of a new species. If the group has moved to an environment that is different from its previous one, natural selection is expected to favor beneficial alleles and eliminate harmful alleles.
Parapatric—two populations occupy overlapping ranges, so a limited amount of interbreeding occurs	A new combination of alleles or chromosomal rearrangement may rapidly limit the amount of gene flow between neighboring populations because hybrid offspring have a very low fitness.
Sympatric—within a population occupying a single habitat in a continuous range, a group evolves into a reproductively isolated species	An abrupt genetic change leads to reproductive isolation. For example, a mutation may affect gamete recognition. In plants, the formation of a tetraploid often leads to the formation of a new species because the interspecies hybrid (e.g., diploid × tetraploid) is triploid and sterile.

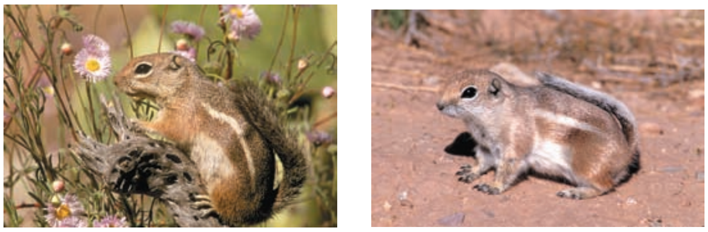

A. harrisi *A. leucurus*

FIGURE 27.3 **An example of allopatric speciation: two closely related species of antelope squirrels that occupy opposite rims of the Grand Canyon.**

Genes → Traits Harris's antelope squirrel (*Ammospermophilus harrisi*) is found on the south rim of the Grand Canyon, whereas the white-tailed antelope squirrel (*Ammospermophilus leucurus*) is found on the north rim. These two species evolved from a common species that existed before the canyon was formed. After the canyon was formed, the two separated populations accumulated genetic changes due to mutation, genetic drift, and natural selection that eventually led to the formation of two distinct species.

accumulation of genetic changes in the two separated populations led to the formation of two morphologically distinct species. Interestingly, birds that can easily fly across the canyon have not diverged into different species on the opposite rims.

Allopatric speciation can also occur via a second mechanism, known as the **founder effect,** which is thought to be more rapid and frequent than allopatric speciation caused by geological events. The founder effect, which was discussed in Chapter 25, occurs when a small group migrates to a new location that is geographically separated from the main population. For example, a storm may force a small group of birds from the mainland to a distant island. In this case, the migration of individuals between the island and the mainland is a very infrequent event. In a relatively short time, the founding population on the island may evolve into a new species. Two evolutionary mechanisms may contribute to this rapid evolution. First, genetic drift may quickly lead to the random fixation of certain alleles and the elimination of other alleles from this smaller population. Another factor is natural selection. The environment on an island may differ significantly from the mainland environment, in which case natural selection on the island may favor certain alleles over others.

Parapatric Speciation **Parapatric speciation** (from the Greek *para*, beside) occurs when members of a species are partially separated. In other words, the geographic separation is not complete. For example, members of a given species may invade a new ecological niche at the periphery of an existing population. Alternatively, a mountain range may divide a species into two populations, but breaks occur in the range where the two groups are connected physically. In these zones of contact, the members of two populations can interbreed, although this tends to occur infrequently. In addition, parapatric speciation may occur among very sedentary species even though no geographic isolation exists. Certain organisms are so sedentary that as little as 100 to 1000 m may be sufficient to limit the interbreeding between neighboring groups. Plants, terrestrial snails, rodents, grasshoppers, lizards, and many flightless insects may speciate in a parapatric manner.

During parapatric speciation, **hybrid zones** exist where two populations can interbreed. For speciation to occur, the amount of gene flow within the hybrid zones must become very limited. In other words, there must be selection against the offspring produced in the hybrid zone. This can happen if each of the two parapatric populations accumulates different chromosomal rearrangements, such as inversions and balanced translocations. How do chromosomal rearrangements, such as inversions, prevent interbreeding? As discussed in Chapter 8, if a hybrid individual has one chromosome with a large inversion and one that does not carry the inversion, crossing over during meiosis can lead to the production of grossly abnormal chromosomes (see Figure 8.10). Therefore, such a hybrid individual is substantially less fertile. By comparison, an individual homozygous for two normal chromosomes or for two chromosomes carrying the same inversion is fertile, because crossing over can proceed normally.

Sympatric Speciation **Sympatric speciation** (from the Greek *sym*, together) occurs when a new species arises in the same geographic area as the species from which it was derived. In plants, a common way for sympatric speciation to occur is the formation of polyploids. As discussed in Chapter 8, complete nondisjunction of chromosomes during gamete formation can increase the number of chromosome sets within a single species (autopolyploidy) or between different species (allopolyploidy) (see Figure 8.25). Polyploidy is so frequent in plants that it is a major form of speciation. In ferns and flowering plants, at least 30% of the species are polyploid. By comparison, polyploidy is much less common in animals, but it can occur. For example, a few dozen species of reptiles and amphibians have been identified that are polyploids derived from diploid relatives.

The formation of a polyploid can abruptly lead to reproductive isolation. As an example, let's consider the probable events that led to the formation of a natural species of common hemp nettle known as *Galeopsis tetrahit*. This species is thought to be an allotetraploid derived from two diploid species: *Galeopsis pubescens* and *Galeopsis speciosa*. As shown in **Figure 27.4a**, *G. tetrahit* has 32 chromosomes, whereas the two diploid species contain 16 chromosomes each ($2n = 16$). **Figure 27.4b** illustrates the chromosomal composition of offspring derived from crosses involving these species. The allotetraploid crossed to another allotetraploid produces an allotetraploid. The allotetraploid is fertile, because all of its chromosomes occur in homologous pairs that can segregate evenly during meiosis. However, a cross between an allotetraploid and a diploid produces an offspring that is monoploid for one chromosome set and diploid for the other. These offspring are expected to be sterile, because they produce highly aneuploid gametes that have incomplete sets of chromosomes. This hybrid sterility renders the allotetraploid reproductively isolated from the diploid species.

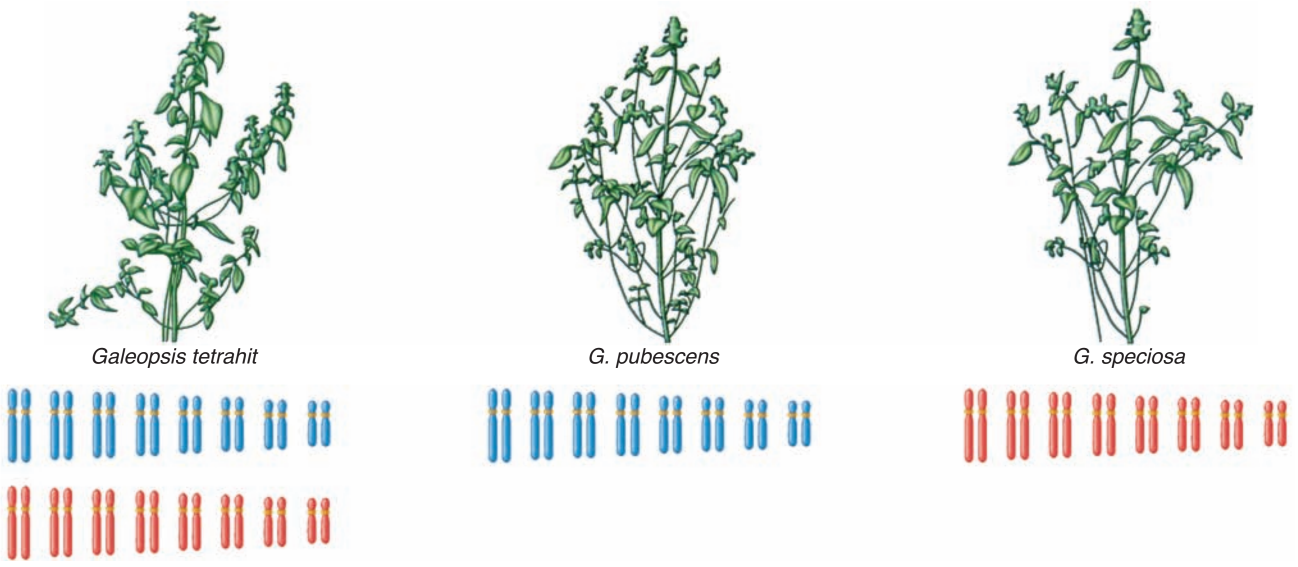
Another way that sympatric speciation may occur is when members of a population occupy different local environments that are continuous with each other. An example of this type of sympatric speciation was described by Jeffrey Feder, Guy Bush, and colleagues. They studied the North American apple maggot fly (*Rhagoletis pomonella*). This fly originally fed on native hawthorn trees. However, the introduction of apple trees approximately 200 years ago provided a new local environment for this species. The apple-feeding populations of this species develop more rapidly because apples mature more quickly than hawthorn fruit. The result is partial temporal isolation in reproduction (see Table 27.1). Although the two populations—those that feed on apple trees and those that feed on hawthorn trees—are considered subspecies, evolutionary biologists speculate they may eventually become distinct species due to reproductive isolation and the accumulation of independent mutations in the two populations.

27.1 REVIEWING THE KEY CONCEPTS

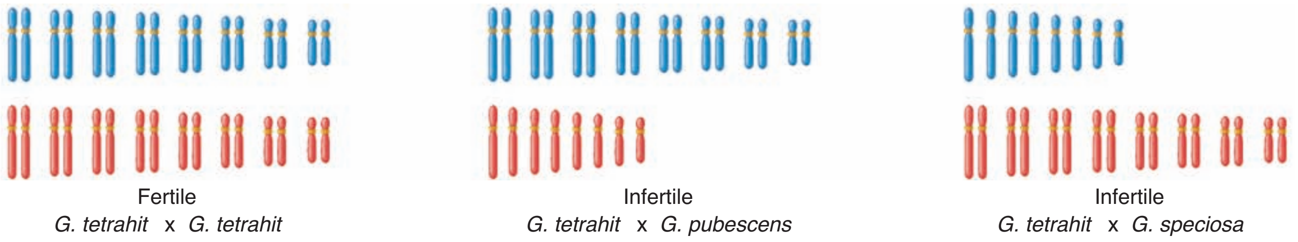
- The most commonly used characteristics for distinguishing species are morphological traits, the ability to interbreed, molecular features, ecological factors, and evolutionary relationships. However, each approach has its drawbacks (see Figure 27.1, Table 27.1).
- A species concept is a way of defining the concept of a species and/or providing an approach for distinguishing one species from another. The general lineage concept is receiving wide support.
- Speciation is the process by which new species are formed via evolution. It may occur by anagenesis, when one species evolves into a new one, or more commonly by cladogenesis, in which a species diverges into two or more different species (see Figure 27.2).
- Cladogenesis may occur via allopatric, parapatric, or sympatric speciation (see Table 27.2, Figures 27.3, 27.4).

27.1 COMPREHENSION QUESTIONS

1. Evolution that results from genotypes with a higher reproductive success is based on
   a. genetic variation.
   b. natural selection.
   c. genetic drift.
   d. both a and b.



(a) Chromosomal composition of 3 *Galeopsis* species


Fertile
G. tetrahit x G. tetrahit
Infertile
G. tetrahit x G. pubescens
Infertile
G. tetrahit x G. speciosa

(b) Outcome of intraspecies and interspecies crosses

FIGURE 27.4 **A comparison of crosses between three natural species of hemp nettle with different ploidy levels.** **(a)** *Galeopsis tetrahit* is an allotetraploid that is thought to be derived from *Galeopsis pubescens* and *Galeopsis speciosa*. **(b)** If *G. tetrahit* is mated with the other two species, the F_1 hybrid offspring will be monoploid for one chromosome set and diploid for the other. The F_1 offspring are likely to be sterile, because they will produce highly aneuploid gametes.

Concept check: *Why is G. tetrahit reproductively isolated from the other two species?*

2. Characteristics that are used to establish species include
   a. morphological traits.
   b. reproductive isolation.
   c. molecular features.
   d. ecological factors.
   e. evolutionary relationships.
   f. all of the above.
3. A pair of birds flies to a deserted island and establishes a colony. Over time, this population evolves into a new species. This is an example of
   a. allopatric speciation.
   b. parapatric speciation.
   c. sympatric speciation.
   d. all of the above.
4. The formation of polyploids is common in plants and can abruptly produce a new species. This is an example of
   a. allopatric speciation.
   b. parapatric speciation.
   c. sympatric speciation.
   d. all of the above.

27.2 PHYLOGENETIC TREES

Thus far, we have considered the various factors that play a role in the formation of new species. In this section, we will examine **phylogeny**—the sequence of events involved in the evolutionary development of a species or group of species. A systematic approach is followed to produce a **phylogenetic tree,** which is a diagram that describes a phylogeny. Such a tree is a hypothesis of the evolutionary relationships among various species, based on the information available to and gathered by biologists termed **systematists.**

Phylogenetic trees are now based on **homology,** which refers to similarities among various species that occur because the species are derived from a common ancestor. Attributes that are the result of homology are said to be **homologous.** For example, the wing of a bat, the arm of a human, and the front leg of a cat are homologous structures. By comparison, a bat wing and insect wing are not homologous; they arose independently of each other. Researchers typically study homology at the level of morphological traits or at the level of genes.

Historically, comparisons of morphological similarities and differences have been used to construct evolutionary trees. In addition, species have been categorized based on physiology, biochemistry, and even behavior. Although these approaches continue to be used, systematists are increasingly using molecular data to infer evolutionary relationships. In 1963, Linus Pauling and Emile Zuckerkandl were the first to suggest the use of molecular data to establish evolutionary relationships. When comparing homologous genes in different species, the DNA sequences from closely related species tend to be more similar to each other than are the sequences from distantly related species. In this section, we will examine the general features of phylogenetic trees, how they can be constructed, and the types of information they reveal.

A Phylogenetic Tree Depicts the Evolutionary Relationships Among Different Species

Let's first take a look at what information is found within a phylogenetic tree and the form in which it is presented. Figure 27.5 shows a hypothetical phylogenetic tree of the relationships between various insect species in which the species (butterflies) are labeled A through J. The vertical axis represents time, with the oldest species at the bottom.

As noted previously, the prevailing mechanism of speciation is through cladogenesis, in which a species diverges into two or more species. The nodes, or branch points, in a phylogenetic tree illustrate times when cladogenesis has occurred. For example, approximately 12 million years ago (mya), species A diverged into species A and species B by cladogenesis. The tips of branches represent species that became extinct in the past, such as species A, B, D, and C, or existing species, such as E, F, G, I, H, and J, that are at the top of the tree.

By studying the branch points of a phylogenetic tree, researchers can group species according to common ancestry. A **monophyletic group,** also known as a **clade,** is a group of species consisting of all descendents of the group's most recent common ancestor. For example, the group highlighted in light green is a clade derived from the common ancestor labeled C. The present-day descendents of a common ancestor can also be called a clade. In this case, species G, I, H, and J form a modern clade. Likewise, the entire tree shown in Figure 27.5 forms a clade, with species A as a common ancestor. As we see in this figure, smaller and more recent clades are subsets of larger ones.

The phylogenetic tree in Figure 27.5 includes ancestral species. This type of tree could be obtained by examining the fossil record. However, many phylogenetic trees do not include ancestral species, but instead focus on the relationships among modern species.

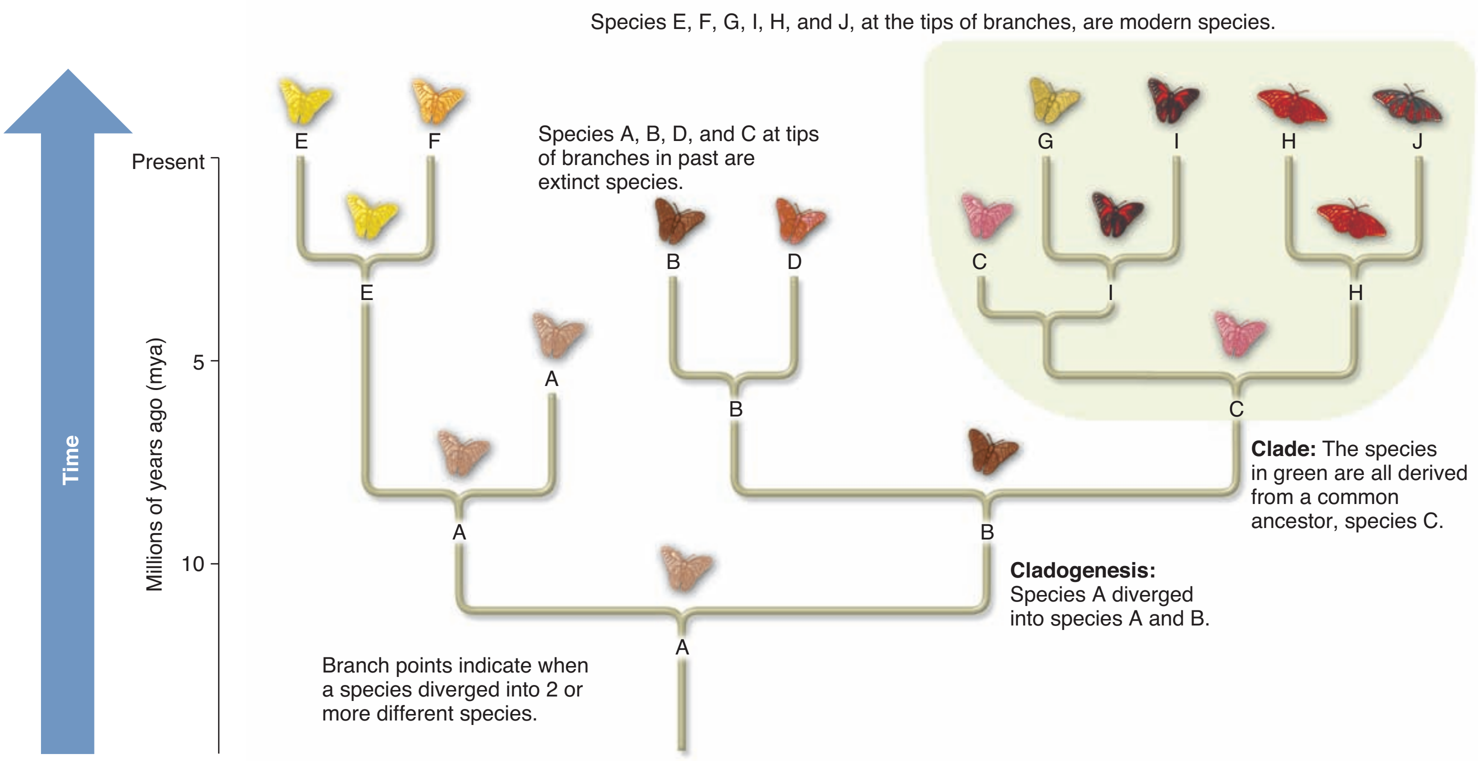

Species E, F, G, I, H, and J, at the tips of branches, are modern species.
Time
Present
Millions of years ago (mya)
5
10
E
F
E
Species A, B, D, and C at tips of branches in past are extinct species.
B
D
A
B
A
B
A
G
I
H
J
C
I
H
C
Clade: The species in green are all derived from a common ancestor, species C.
Cladogenesis: Species A diverged into species A and B.
Branch points indicate when a species diverged into 2 or more different species.

FIGURE 27.5 **How to read a phylogenetic tree.** This hypothetical tree shows the proposed relationships among various butterfly species.

Concept check: *What is a clade?*

A Phylogenetic Tree Can Be Constructed Using Cladistic and Phenetic Approaches

Now that we appreciate the concept of a phylogenetic tree, let's turn our attention to how evolutionary biologists actually construct them.

The most popular methods of building phylogenetic trees can be classified into two broad categories called phenetics and cladistics. A **phenetic approach** constructs a phylogenetic tree by considering the overall similarities among a group of species without trying to understand their evolutionary history. Such trees are called **phenograms.** By comparison, in a **cladistic approach,** a tree is constructed by considering the various possible pathways of evolution and choosing from among the best possible trees. In this approach, a goal is to group species based on knowledge concerning traits that occurred earlier in evolution versus traits that came later. Such trees are called **cladograms.**

Which approach is better? For data involving morphological traits and for the construction of complex evolutionary trees that include many taxonomic levels, the cladistic approach is generally superior. Even so, both approaches are used, and many evolutionary biologists construct their evolutionary trees based on a combination of both cladistics and phenetics. We will briefly consider both types of methods next.

Cladistics In the 1950s, the German entomologist Willi Hennig proposed that evolutionary relationships should be inferred from new features shared by descendants of a common ancestor. A cladistic approach compares traits, also called characters, that are either shared or not shared by different species. These can be morphological traits, such as the shapes of birds' beaks, or molecular traits, such as sequences of homologous genes. Such characters may come in different versions called character states. Those that are shared with a distant ancestor are called **ancestral characters** (also called **primitive characters**). Such characters are viewed as being older—ones that arose earlier in evolution. In contrast, a **shared derived character,** or **synapomorphy,** is a trait that is shared by a group of organisms but not by a distant common ancestor. Compared with ancestral characters, shared derived characters are more recent traits on an evolutionary timescale. For example, among mammals, only some species have flippers, such as whales and dolphins. In this case, flippers were derived from the two front limbs of an ancestral species. The word *derived* refers to the observation that evolution involves the modification of traits in preexisting species. In other words, the features of newer populations of organisms are derived from changes in preexisting populations. The basis of the cladistic approach is to analyze many shared derived characters among groups of species to deduce the pathway that gave rise to the species.

To understand the concept of ancestral versus shared derived characters, let's consider how a cladogram can be constructed based on molecular data such as a sequence of a gene. Our example uses molecular data obtained from seven different hypothetical species called A through G. In these species, the same gene was sequenced from seven different species; a portion of the gene sequence is shown as follows:

A: GATAGTACCC
B: GATAGTTCCC
C: GATAGTTCCG
D: GGTATTACCC
E: GGTATAACCC
F: GGTAGTACCA
G: GGTAGTACCC

In a cladogram, an **ingroup** is a monophyletic group in which we are interested. By comparison, an **outgroup** is a species or group of species that is more distantly related to an ingroup. The root of a cladogram is placed between the outgroup and the ingroup. In the cladogram of Figure 27.6, the outgroup is species E. This may have been inferred because the other species may share traits that are not found in species E. The other species (A,

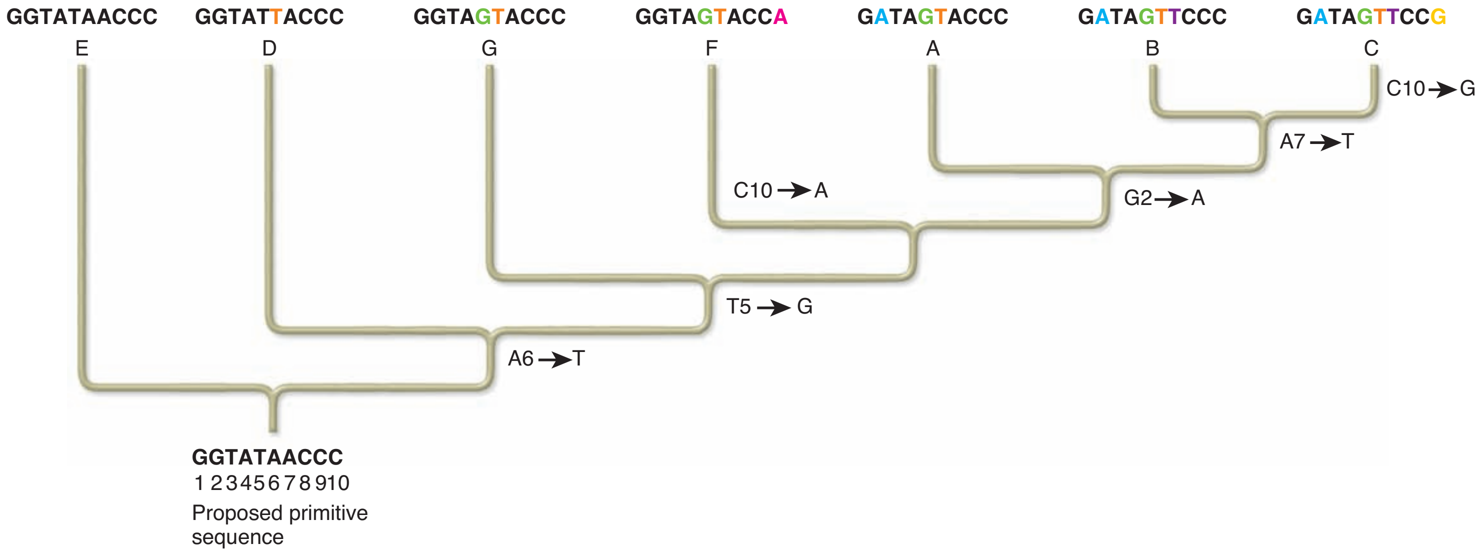


FIGURE 27.6 **Shared derived characters involving a molecular trait.** This phylogenetic tree illustrates a cladogram of relationships involving homologous gene sequences found in seven species. Mutations that alter an ancestral DNA sequence are shared among certain species, thereby allowing the construction of a cladogram.

Concept check: *What is the difference between an ancestral character and a shared derived character?*

B, C, D, F, and G) form the ingroup. For these data, a mutation that changes the DNA sequence is analogous to a modification of a characteristic. Species that share such genetic changes possess shared derived characters because the new genetic sequence was derived from a more ancestral sequence.

Now that we understand some of the general principles of cladistics, let's consider the steps a researcher follows to construct a cladogram using this approach.

1. **Choose the species in whose evolutionary relationships you are interested.** In a simple cladogram, individual species are compared with each other. In more complex cladograms, species may be grouped into larger taxa (e.g., families) and compared with each other. If such grouping is done, the groups must be clades for the results to be reliable.
2. **Choose characters for comparing different species.** As mentioned, a character is a general feature of an organism. Characters may come in different versions called character states. For example, a base at a particular location in a gene can be considered a character, and this character could exist in different character states, such as A, T, G, or C, due to mutations.
3. **Determine the polarity of character states.** In other words, determine if a character state is ancestral or derived. In the case of morphological traits, this information may be available by examining the fossil record. If possible, identify an outgroup.
4. **Group species (or higher taxa) based on shared derived characters.**
5. **Build a cladogram based on the following principles:**
    - All species (or higher taxa) are placed on tips in the phylogenetic tree, not at branch points. A cladogram does not include ancestral species at branch points.
    - Each cladogram branch point should have a list of one or more shared derived characters that are common to all species above the branch point unless the character is later modified.
    - All shared derived characters appear together only once in a cladogram unless they independently arose during evolution more than once in the ancestors of different clades.
6. **Choose the best cladogram among possible options.** When grouping species (or higher taxa), more than one cladogram may be possible. Therefore, analyzing the data and producing the best possible cladogram is a key aspect of this process. As described next, different approaches can be followed to achieve this goal.

The greatest challenge in a cladistic approach is to determine the correct order of events. It may not always be obvious which traits are ancestral and came earlier and which are derived and came later in evolution. Different approaches can be used to deduce the correct order. First, for morphological traits, a common way to deduce the order of events is to analyze fossils and determine the relative dates that certain traits arose. A second strategy assumes that the best hypothesis is the one that requires the fewest number of evolutionary changes. This concept, called the **principle of parsimony,** states that the preferred hypothesis is the one that is the simplest. For example, if two species possess a tail, we would initially assume that a tail arose once during evolution and that both species have descended from a common ancestor with a tail. Such a hypothesis is simpler than assuming that tails arose twice during evolution and that the tails in the two species are not due to descent from a common ancestor.

Let's consider a simple example to illustrate how the principle of parsimony can be used. Our example involves molecular data obtained from four different hypothetical species called A through D. In these species, a homologous region of DNA was sequenced as shown here.

```
   12345
A  GTACA
B: GACAG
C: GTCAA
D: GACCG
```

Given this information, three different trees are shown in Figure 27.7, although more are possible. In these examples, tree 1 requires seven mutations, tree 2 requires six, and tree 3 requires only five. Because tree 3 requires the fewest number of mutations, it is considered the most parsimonious. Based on the principle of parsimony, tree 3 is the more likely choice.

Phenetics As mentioned, a phenetic approach constructs a phylogenetic tree based on overall similarities among a group of species. A relatively simple method is the UPGMA method (unweighted pair group method with arithmetic mean). When comparing gene sequences, this method assumes that after two species have diverged from each other, the rate of accumulation of neutral mutations in a homologous region of DNA is approximately the same in the two species. As discussed later in Section 27.3, this means that the molecular clock for neutral mutations is linear. If this assumption is met, closely related sequences (i.e., those that have diverged more recently) have greater sequence similarity compared to distantly related species.

As an example of the UPGMA method, we can consider the nucleotide similarities among five species of primates. In a homologous region containing 10,000 bp, the following numbers of nucleotide substitutions were found:

	Human	Chimpanzee	Gorilla	Orangutan	Rhesus monkey
Human	0	145	151	298	751
Chimpanzee	145	0	157	294	755
Gorilla	151	157	0	304	739
Orangutan	298	294	304	0	710
Rhesus monkey	751	755	739	710	0

To construct a tree, we begin with the pair that is the most closely related. In this case, it would be humans and chimpanzees, because that pair has the fewest number of nucleotide substitutions. The number of nucleotide substitutions per 10,000 nucleotides equals 145. We divide 145 by 2 to yield the average number of substitutions, which occurred in each species since the time they diverged. This number equals 72.5. In other words, about 72.5 substitutions occurred in humans and about 72.5 (different) substitutions occurred in chimpanzees since they diverged from a common ancestor and that is why there are 145 nucleotide differences today. It is common to express this value as the percentage of nucleotide differences. So,

$$\underset{\text{(humans vs chimpanzee)}}{\text{Percentage of nucleotide differences}} = \frac{72.5 \times 100}{10{,}000} = 0.725$$

Next, we consider the species that is the most closely related to humans and chimpanzees. That would be the gorilla. There are 151 substitutions between gorillas and humans, and 157 substitutions between gorillas and chimpanzees. The average number of substitutions between gorillas and humans/chimpanzees equals 151 + 157 divided by 2, which equals 154. We divide 154 by 2 to yield the average number of substitutions, which occurred in each species since the time they diverged from a common ancestor. This number equals 77. In other words, about 77 substitutions occurred in humans or chimpanzees and about 77 (different) substitutions occurred in gorillas since they diverged from a common ancestor and that is why there are 154 nucleotide differences today. To express this value as the percentage of nucleotide differences:

$$\underset{\text{(gorillas vs humans/chimpanzees)}}{\text{Percentage of nucleotide substitutions}} = \frac{77 \times 100}{10{,}000} = 0.77$$

Next, we consider the species that is the most closely related to humans, chimpanzees, and gorillas. That would be orangutans. There are 298 substitutions between orangutans and humans, 294 substitutions between orangutans and chimpanzees, and 304 substitutions between orangutans and gorillas. The average number of substitutions between orangutans and humans/chimpanzees/gorillas equals 298 + 294 + 304 divided by 3, which equals 299. We divide 299 by 2 to yield the average number of substitutions, which occurred in each species since the time, they diverged from a common ancestor. This number approximately equals 149. To express this value as the percentage of nucleotide differences:

$$\underset{\text{(orangutans vs humans/chimpanzees/gorillas)}}{\text{Percentage of nucleotide substitutions}} = \frac{149 \times 100}{10{,}000} = 1.49$$

Finally, we consider the most distantly related species, which is the Rhesus monkey. There are 751 substitutions between Rhesus monkeys and humans, 755 substitutions between Rhesus monkeys and chimpanzees, 739 substitutions between Rhesus monkeys and gorillas, and 710 substitutions between Rhesus monkeys and orangutans. The average number of substitutions between Rhesus monkeys compared with the humans, chimpanzees, gorillas, or orangutans equals 751 + 755 + 739 + 710 divided by 4, which equals 739. We divide 739 by 2 to yield the average number of substitutions, which occurred in each species since the time they diverged from a common ancestor. This number approximately equals 369. To express this value as the percentage of nucleotide differences:

$$\underset{\substack{\text{(Rhesus monkeys vs humans/chimpanzees/)}\\ \text{gorillas/orangutans)}}}{\text{Percentage of nucleotide substitutions}} = \frac{369 \times 100}{10{,}000} = 3.69$$

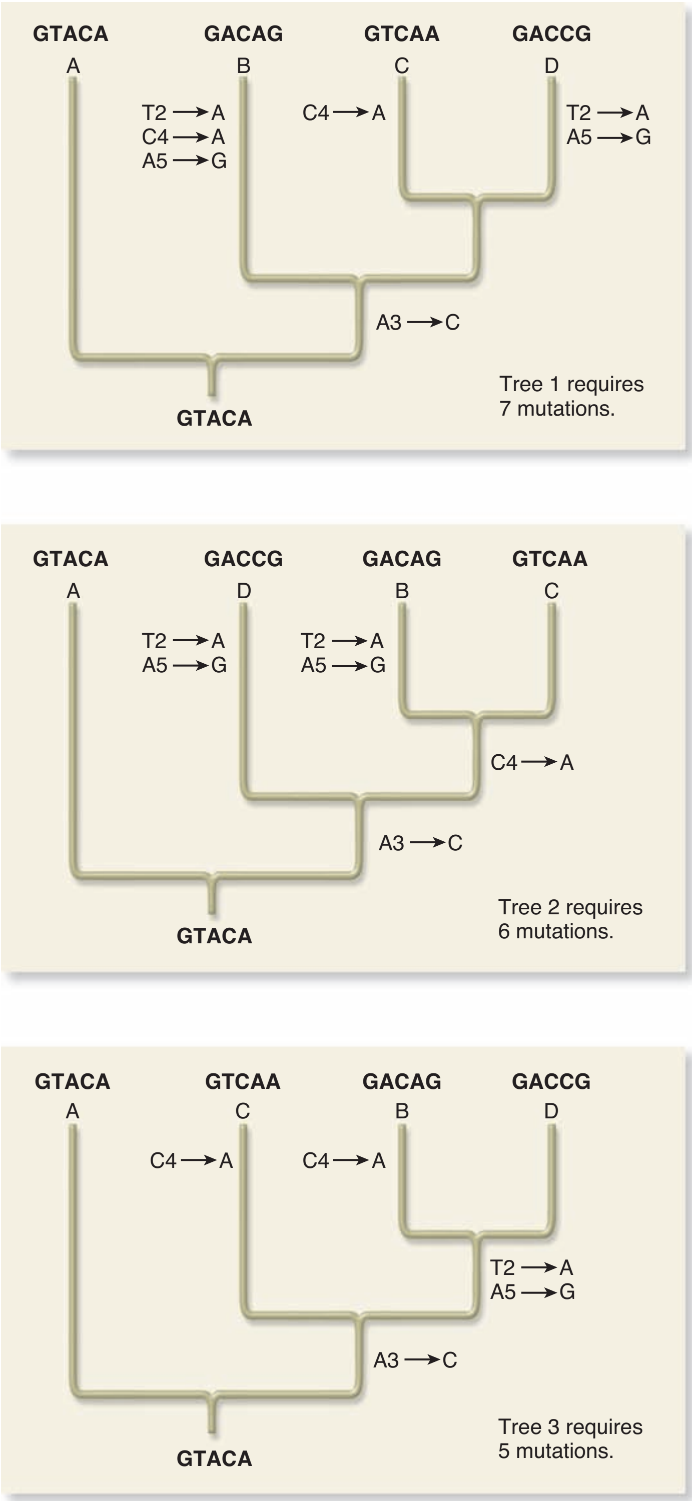


FIGURE 27.7 **The cladistic approach: Choosing a cladogram from molecular genetic data.** This figure shows three different phylogenetic trees for the evolution of a short DNA sequence, but many more are possible. According to the principle of parsimony, the cladogram shown in tree 3 is the more plausible choice because it requires only five mutations. When constructing cladograms based on long genetic sequences, researchers use computers to generate trees with the fewest possible genetic changes.

Concept check: *What is the principle of parsimony?*

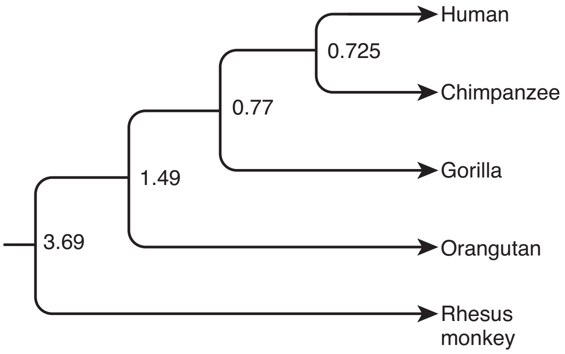


FIGURE 27.8 **Reconstruction of a phylogenetic tree using the UPGMA method.**

With these data, we can construct the phylogenetic tree shown in Figure 27.8. This illustration shows the relative evolutionary relationships among these primate species.

Maximum Likelihood and Bayesian Methods Are Used to Discriminate Among Possible Phylogenetic Trees

Evolutionary biologists also apply other approaches, such as maximum likelihood and Bayesian methods, when proposing and evaluating phylogenetic trees. These methods involve the use of an evolutionary model—a set of assumptions about how evolution is likely to happen. For example, mutations affecting the third base in a codon are often neutral because they don't affect the amino acid sequence of the encoded protein and therefore don't affect the fitness of an organism. As discussed later in this chapter, such neutral mutations are more likely to become prevalent in a population than are mutations in the first or second base. Therefore, one possible assumption of an evolutionary model is that neutral mutations are more likely than nonneutral ones.

According to an approach called **maximum likelihood,** researchers ask the question: What is the probability that an evolutionary model and a proposed phylogenetic tree would give rise to the observed data? The rationale is that a phylogenetic tree that gives a higher probability of producing the observed data is preferred to any trees that give a lower probability. By comparison, **Bayesian methods** ask the question: What is the probability that a particular phylogenetic tree is correct, given the observed data and a particular evolutionary model? Though the computational strategies of maximum likelihood and Bayesian methods are different (and beyond the scope of this textbook), the goal of both approaches is to identify one or more trees that are most likely to be correct based on an evolutionary model and the available data.

Phylogenetic Trees Refine Our Understanding of Evolutionary Relationships

For molecular evolutionary studies, the DNA sequences of many genes have been obtained from a wide range of sources. Several different types of gene sequences have been used to construct phylogenetic trees. One very commonly analyzed gene is that encoding 16S rRNA (bacterial) or 18S rRNA (eukaryotic), an rRNA found in the small ribosomal subunit. This gene has been sequenced from thousands of different species. Because rRNA is universal in all living organisms, its function was established at an early stage in the evolution of life on this planet, and its sequence has changed fairly slowly. Presumably, most mutations in this gene are deleterious, so few neutral or beneficial alleles can occur. This limitation causes this gene sequence to change very slowly during evolution. Furthermore, 16S and 18S rRNAs are rather large molecules and therefore contain a large amount of sequence information.

In 1977, Carl Woese analyzed 16S and 18S rRNA sequences and identified a new domain of life called Archaea. His work proposed three main evolutionary branches: the **Bacteria,** the **Archaea,** and the **Eukaryotes** (also called Eukarya). From these types of genetic analyses, it has become apparent that all living organisms are connected through a complex evolutionary tree.

Although the work of Woese was a breakthrough in our appreciation of evolution, more recent molecular genetic data have shed new light on the classification of species. Specifically, biologists once categorized eukaryotic species into four kingdoms: protists (Protista), fungi (Fungi), animals (Animalia), and plants (Plantae). However, recent models propose several major groups, called **supergroups,** as a way to organize eukaryotes into monophyletic groups. Figure 27.9 shows a diagram that hypothesizes several supergroups. Each supergroup contains distinctive types of protists. In addition, Fungi and Animalia are within the Opisthokonta supergroup, and Plantae is found within Land plants and algal relatives. As seen in this figure, molecular data and newer ways of building trees reveal that protists played a key role in the evolution of many diverse groups of eukaryotic species, producing several large monophyletic supergroups.

Horizontal Gene Transfer Also Contributes to the Evolution of Species

The types of phylogenetic trees considered thus far are examples of **vertical evolution,** in which species evolve from preexisting species by the accumulation of gene mutations and by changes in chromosome structure and number. Vertical evolution involves genetic changes in a series of ancestors that form a lineage. In addition to vertical evolution, however, species accumulate genetic changes by another process called **horizontal gene transfer.** This refers to a process in which an organism incorporates genetic material from another organism without being the offspring of that organism. It often involves the exchange of genetic material between different species (see Chapter 25, Figure 25.20).

An analysis of many genomes suggests that horizontal gene transfer was prevalent during the early stages of evolution, when all organisms were unicellular, but continued even after the divergence of the three major domains of life. With regard to modern organisms, horizontal gene transfer remains prevalent among

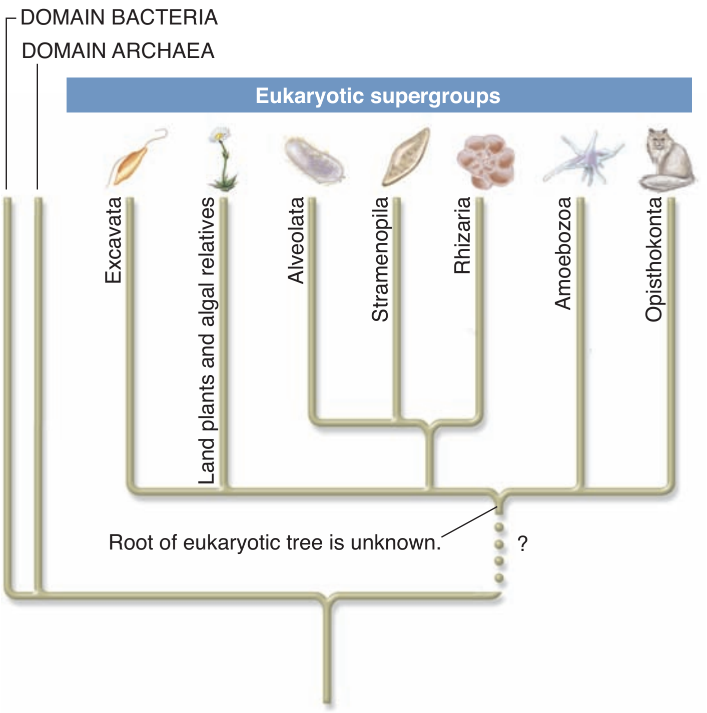


FIGURE 27.9 **A modern cladogram for eukaryotes.** Each of the supergroups shown here are hypothesized to be monophyletic. This drawing should be considered a working hypothesis. The arrangement of these supergroups relative to each other is not entirely certain.

Concept check: Discuss where protists are found in this newer grouping of eukaryotic species.

prokaryotic species. By comparison, this process is less common among eukaryotes, though it does occur. Researchers have speculated that multicellularity and sexual reproduction have presented barriers to horizontal gene transfer in many eukaryotes. For a gene to be transmitted to eukaryotic offspring, it must be transferred into a eukaryotic gamete or a cell that gives rise to gametes.

How has horizontal gene transfer affected evolution? In the past few decades, scientists have debated the role of horizontal gene transfer in the earliest stages of evolution, prior to the emergence of the two prokaryotic domains. The traditional viewpoint was that the three domains of life arose from a single type of prokaryotic cell (or pre-prokaryotic cell) called the universal ancestor. However, genomic research has suggested that horizontal gene transfer may have been particularly common during the early stages of evolution on Earth, when all species were unicellular. Rather than proposing that all life arose from a single type of prokaryotic cell, horizontal gene transfer may have been so prevalent that the universal ancestor may have actually been an ancestral community of cell lineages that evolved as a whole. If that were the case, the tree of life cannot be traced back to a single universal ancestor.

Figure 27.10 illustrates a schematic scenario for the evolution of life on Earth that includes the roles of both vertical evolution and horizontal gene transfer. This has been described as a "web of life" rather than a "tree of life." Instead of a universal ancestor, a web of life began with a community of primitive cells that transferred genetic material in a horizontal fashion. Horizontal gene transfer was also prevalent during the early evolution of bacteria and archaea and when eukaryotes first emerged as unicellular species. Although horizontal gene transfer remains a prominent way to foster evolutionary change in modern bacteria and archaea, the region of the diagram that contains eukaryotic species has a more treelike structure, because horizontal gene transfer has become much less common in these species.

27.2 REVIEWING THE KEY CONCEPTS

- A phylogenetic tree is a diagram that describes a phylogeny—the sequence of events involved in the evolutionary development of a species or group of species (see Figure 27.5).
- Different methods, including cladistics and phenetics, are used to construct phylogenetic trees. One way to evaluate the validity of possible trees is by the principle of parsimony (see Figures 27.6, 27.7, 27.8).
- Maximum likelihood and Bayesian methods are also used to discriminate among possible phylogenetic trees.
- Phylogenetic trees help us to understand the relationships among different taxa such as domains and supergroups (see Figure 27.9).
- In addition to vertical evolution, horizontal gene transfer has played an important role in the evolution of life on Earth (see Figure 27.10).

27.2 COMPREHENSION QUESTIONS

1. Phylogenetic trees are based on
   a. natural selection.
   b. genetic drift.
   c. homology.
   d. none of the above.
2. A shared derived character is
   a. derived from an ancestral character.
   b. more recent compared with an ancestral character.
   c. found in all or most of the members of an ingroup.
   d. all of above.
3. An approach that is used to propose a phylogenetic tree is
   a. cladistics and the principle of parsimony.
   b. phenetics.
   c. maximum likelihood and Bayesian methods.
   d. all of the above.
4. Horizontal gene transfer is a process in which genetic material from an organism is
   a. transferred from cell to cell.
   b. transferred to its offspring.
   c. transferred to another organism that is not the offspring of that organism.
   d. released into the environment.

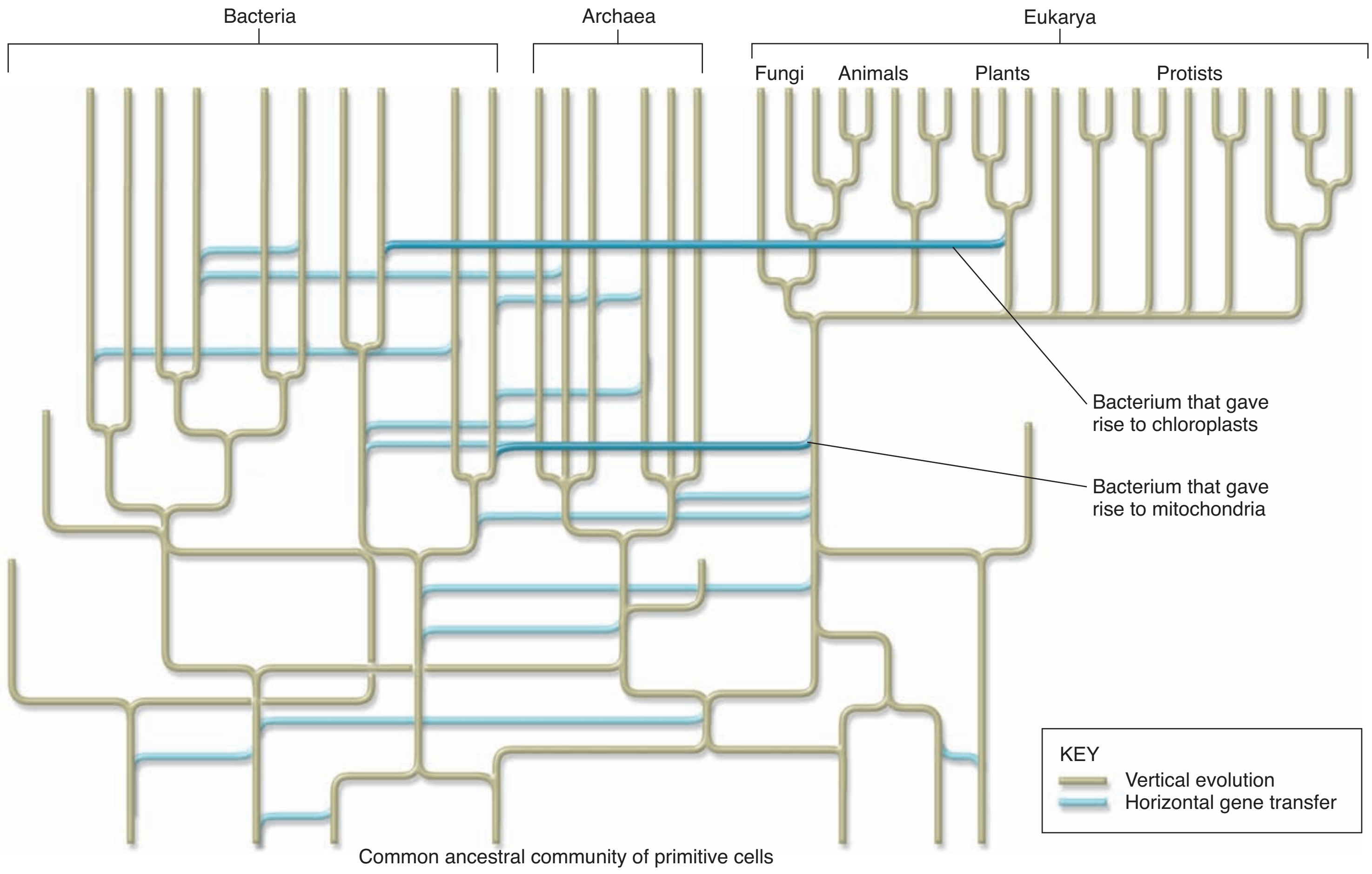

Bacteria
Archaea
Eukarya
Fungi
Animals
Plants
Protists
Bacterium that gave rise to chloroplasts
Bacterium that gave rise to mitochondria
KEY
Vertical evolution
Horizontal gene transfer
Common ancestral community of primitive cells

FIGURE 27.10 **A revised view regarding the evolution of life, incorporating the concept of horizontal gene transfer.** This phylogenetic tree shows a classification of life on Earth that includes the contribution of horizontal gene transfer in the evolution of species on our planet. This phenomenon was prevalent during the early stages of evolution when all organisms were unicellular. Horizontal gene transfer continues to be a prominent factor in the speciation of bacteria and archaea. Note: This tree is meant to be schematic. Figure 27.9 is a more realistic representation of the evolutionary relationships among modern species. In this simplified drawing, the horizontal gene transfer that gave rise to chloroplasts is shown as a single event, though multiple transfers of plastids have occurred during the evolution of life on Earth.

Concept check: *What is horizontal gene transfer?*

27.3 MOLECULAR EVOLUTION AND MOLECULAR CLOCKS

Molecular evolution refers to patterns and processes associated with evolutionary change at the molecular level. Such changes may be phenotypically neutral, or they may underlie the phenotypic changes associated with evolution. Evolutionary changes at the DNA level can be objectively compared among different species to establish evolutionary relationships. Furthermore, this approach can be used to compare any two existing organisms, no matter how greatly they differ in their morphological traits. For example, we can compare DNA sequences between humans and bacteria, or between plants and fruit flies. Differences in DNA sequences are quantitative and can be analyzed using computer programs. Such comparisons would be very difficult at a morphological level. In this section, we will examine how evolution occurs at the molecular level.

Homologous Genes Are Derived from a Common Ancestral Gene

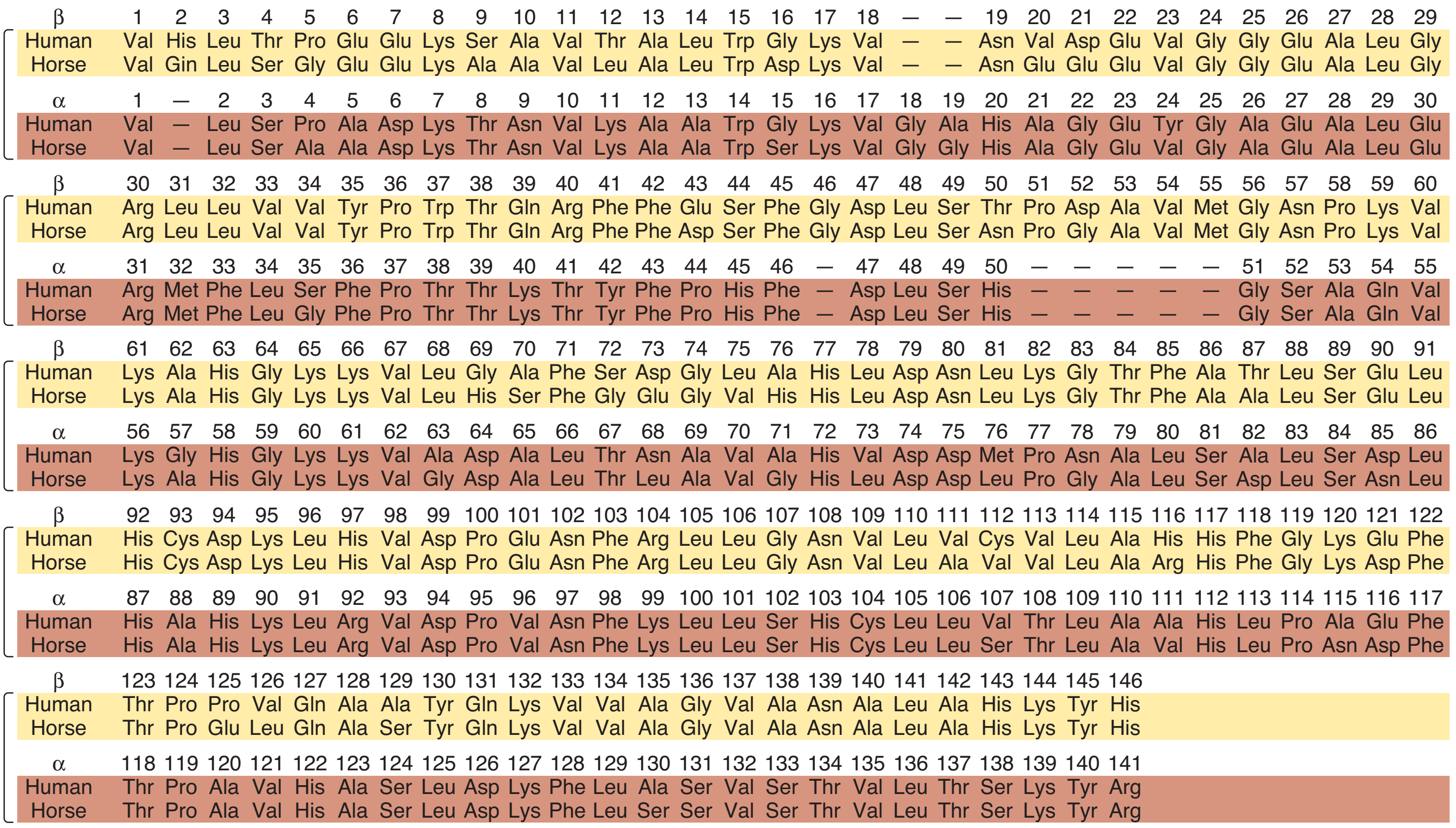
As we have seen, a phylogenetic tree is based on homology, which refers to similarities among various species that occur because the species are derived from a common ancestor. In this section, we will focus on genetic homology. Two genes are said to be homologous if they are derived from the same ancestral gene. During cladogenesis, a single species diverges into two or more different species. When two homologous genes are found in different species, these genes are termed **orthologs.** In Chapter 24, we considered orthologs of *Hox* genes. In that case, several homologous genes were identified in the fruit fly and the mouse. In addition, two or more homologous genes can be found within a single species. These are termed paralogous genes, or **paralogs.** As discussed in Chapter 8, this occurs when abnormal gene duplication events produce multiple copies of a gene, which ultimately leads to the formation of a **gene**

family. A gene family consists of two or more paralogs within the genome of a particular species.

Figure 27.11a shows examples of both orthologs and paralogs in the globin gene family. Hemoglobin is an oxygen-carrying protein found in all vertebrate species. It is composed of two different subunits, encoded by the α-globin and β-globin genes. Figure 27.11a shows an alignment of the deduced amino acid sequences encoded by these genes. The sequences are homologous between humans and horses because of their evolutionary relationship. We say that the human and horse α-globin genes are orthologs of each other, as are the human and horse β-globin genes. The α-globin and β-globin genes in humans are paralogs of each other.

As shown in **Figure 27.11b**, the sequences of the orthologs are more similar to each other than they are to the paralogs. For example, the sequences of human and horse β globins show 25 differences, whereas the human β globin and human α globins show 84 differences. What do these results mean? They indicate that the gene duplication that led to the formation of the α-globin and β-globin genes occurred long before the evolutionary divergence that produced different species of mammals. For this reason, there was a greater amount of time for the α- and β-globin genes to accumulate changes compared with the amount of time that has elapsed since the evolutionary divergence of mammalian species. This idea is schematically shown in **Figure 27.12**.

Based on the analysis of genetic sequences, evolutionary biologists have estimated that the gene duplication that produced the α-globin and β-globin gene lineages occurred approximately 400 million years ago (mya), whereas the speciation events that resulted in different species of mammals occurred less than 200 mya. Therefore, the α-globin and β-globin genes have had much more time to accumulate changes relative to each other.



Hemoglobin

β	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	—	—	19	20	21	22	23	24	25	26	27	28	29
Human	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	Ser	Ala	Val	Thr	Ala	Leu	Trp	Gly	Lys	Val	—	—	Asn	Val	Asp	Glu	Val	Gly	Gly	Glu	Ala	Leu	Gly
Horse	Val	Gin	Leu	Ser	Gly	Glu	Glu	Lys	Ala	Ala	Val	Leu	Ala	Leu	Trp	Asp	Lys	Val	—	—	Asn	Glu	Glu	Glu	Val	Gly	Gly	Glu	Ala	Leu	Gly
α	1	—	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Human	Val	—	Leu	Ser	Pro	Ala	Asp	Lys	Thr	Asn	Val	Lys	Ala	Ala	Trp	Gly	Lys	Val	Gly	Ala	His	Ala	Gly	Glu	Tyr	Gly	Ala	Glu	Ala	Leu	Glu
Horse	Val	—	Leu	Ser	Ala	Ala	Asp	Lys	Thr	Asn	Val	Lys	Ala	Ala	Trp	Ser	Lys	Val	Gly	Gly	His	Ala	Gly	Glu	Val	Gly	Ala	Glu	Ala	Leu	Glu

β	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Human	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe	Gly	Asp	Leu	Ser	Thr	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val
Horse	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Asp	Ser	Phe	Gly	Asp	Leu	Ser	Asn	Pro	Gly	Ala	Val	Met	Gly	Asn	Pro	Lys	Val
α	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	—	47	48	49	50	—	—	—	—	—	51	52	53	54	55
Human	Arg	Met	Phe	Leu	Ser	Phe	Pro	Thr	Thr	Lys	Thr	Tyr	Phe	Pro	His	Phe	—	Asp	Leu	Ser	His	—	—	—	—	—	Gly	Ser	Ala	Gln	Val
Horse	Arg	Met	Phe	Leu	Gly	Phe	Pro	Thr	Thr	Lys	Thr	Tyr	Phe	Pro	His	Phe	—	Asp	Leu	Ser	His	—	—	—	—	—	Gly	Ser	Ala	Gln	Val

β	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Human	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu	Ala	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Thr	Leu	Ser	Glu	Leu
Horse	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	His	Ser	Phe	Gly	Glu	Gly	Val	His	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Ala	Leu	Ser	Glu	Leu
α	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86
Human	Lys	Gly	His	Gly	Lys	Lys	Val	Ala	Asp	Ala	Leu	Thr	Asn	Ala	Val	Ala	His	Val	Asp	Asp	Met	Pro	Asn	Ala	Leu	Ser	Ala	Leu	Ser	Asp	Leu
Horse	Lys	Ala	His	Gly	Lys	Lys	Val	Gly	Asp	Ala	Leu	Thr	Leu	Ala	Val	Gly	His	Leu	Asp	Asp	Leu	Pro	Gly	Ala	Leu	Ser	Asp	Leu	Ser	Asn	Leu

β	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122
Human	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn	Phe	Arg	Leu	Leu	Gly	Asn	Val	Leu	Val	Cys	Val	Leu	Ala	His	His	Phe	Gly	Lys	Glu	Phe
Horse	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn	Phe	Arg	Leu	Leu	Gly	Asn	Val	Leu	Ala	Val	Val	Leu	Ala	Arg	His	Phe	Gly	Lys	Asp	Phe
α	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117
Human	His	Ala	His	Lys	Leu	Arg	Val	Asp	Pro	Val	Asn	Phe	Lys	Leu	Leu	Ser	His	Cys	Leu	Leu	Val	Thr	Leu	Ala	Ala	His	Leu	Pro	Ala	Glu	Phe
Horse	His	Ala	His	Lys	Leu	Arg	Val	Asp	Pro	Val	Asn	Phe	Lys	Leu	Leu	Ser	His	Cys	Leu	Leu	Ser	Thr	Leu	Ala	Val	His	Leu	Pro	Asn	Asp	Phe

β	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146
Human	Thr	Pro	Pro	Val	Gln	Ala	Ala	Tyr	Gln	Lys	Val	Val	Ala	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His
Horse	Thr	Pro	Glu	Leu	Gln	Ala	Ser	Tyr	Gln	Lys	Val	Val	Ala	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His
α	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141
Human	Thr	Pro	Ala	Val	His	Ala	Ser	Leu	Asp	Lys	Phe	Leu	Ala	Ser	Val	Ser	Thr	Val	Leu	Thr	Ser	Lys	Tyr	Arg
Horse	Thr	Pro	Ala	Val	His	Ala	Ser	Leu	Asp	Lys	Phe	Leu	Ser	Ser	Val	Ser	Thr	Val	Leu	Thr	Ser	Lys	Tyr	Arg

(a) Alignment of globin genes

Orthologs	**Number of Amino Acid Differences**
Human β globin vs horse β globin	25 out of 146
Human α globin vs horse α globin	18 out of 141
Paralogs	
Human β globin vs human α globin	84 out of 146
Horse β globin vs horse α globin	81 out of 146

(b) A comparison of amino acid differences between orthologs and paralogs

FIGURE 27.11 **A comparison of the α- and β-globin polypeptides from humans and horses.** **(a)** An alignment of the deduced amino acid sequences obtained by sequencing the exon portions of the corresponding genes. The gaps indicate where additional amino acids are found in the sequence of myoglobin, another member of this gene family. **(b)** A comparison of amino acid differences between orthologs and paralogs.

Concept check: *What is the difference between a paralog and an ortholog?*

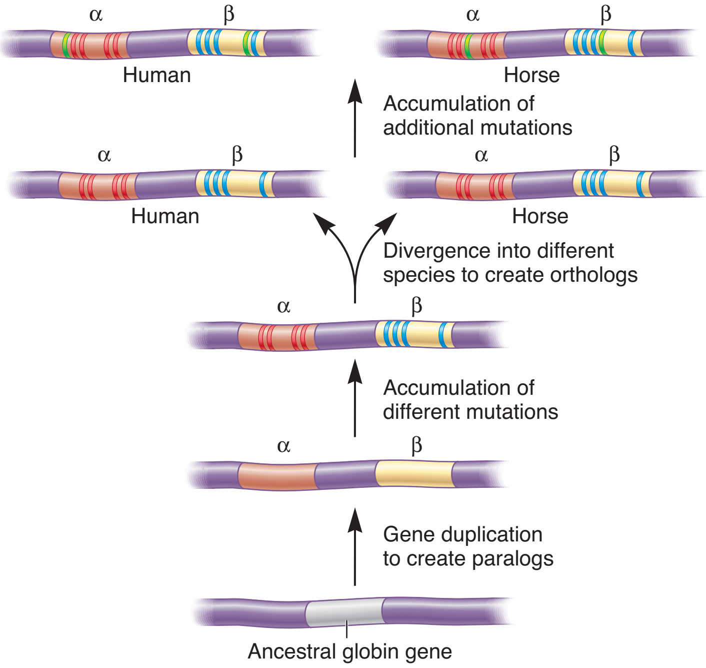

α
β
Human
α
β
Horse
Accumulation of additional mutations
α
β
Human
α
β
Horse
Divergence into different species to create orthologs
α
β
Accumulation of different mutations
α
β
Gene duplication to create paralogs
Ancestral globin gene

FIGURE 27.12 **Evolution of paralogs and orthologs.** In this schematic example, the ancestral globin gene duplicated to create the α- and β-globin genes, which are paralogous genes. (Note: Actually, gene duplications occurred several times to produce a large globin gene family, but this is a simplified example that shows only one gene duplication event. Also, chromosomal rearrangements have placed the α- and β-globin genes on different chromosomes.) Over time, these two paralogs accumulated different mutations, designated with red and blue lines. At a later point in evolution, a species divergence occurred to create different mammalian species including humans and horses. Over time, the orthologs also accumulated different mutations, designated with green lines. As noted here, the orthologs have fewer differences than the paralogs because the gene duplication occurred prior to the species divergence.

Genetic Variation at the Molecular Level Is Associated with Neutral Changes in Gene Sequences

As we have seen, the globin genes exhibit variation in their sequences. How does this variation arise? Is it due primarily to mutations that are favored by natural selection or due primarily to random genetic drift? This has been the subject of debate by biologists.

A **nonneutral mutation** is one that affects the phenotype of the organism and can be acted on by natural selection. Such a mutation may only subtly alter the phenotype of an organism, or it may have a major effect. According to Darwin, natural selection is the agent that leads to evolutionary change in populations. It selects for individuals with the highest Darwinian fitness and often promotes the establishment of beneficial alleles and the elimination of deleterious ones. Therefore, many geneticists have assumed that natural selection is the dominant factor in changing the genetic composition of natural populations, thereby leading to variation.

In opposition to this viewpoint, in 1968, Motoo Kimura proposed the **neutral theory of evolution.** According to this theory, most genetic variation observed in natural populations is due to the accumulation of neutral mutations that do not affect the phenotype of the organism and are not acted on by natural selection. For example, a mutation within a structural gene that changes a glycine codon from GGG to GGC does not affect the amino acid sequence of the encoded protein. Because neutral mutations do not affect phenotype and do not provide the individual with a selective advantage, they spread throughout a population according to their frequency of appearance and to random genetic drift. This theory has been called the "survival of the luckiest" to contrast it with Darwin's theory, which focuses on fitness. Kimura did not dispute that natural selection is responsible for adaptive changes in a species during evolution. His main argument is that most modern variation in gene sequences is neutral with respect to natural selection.

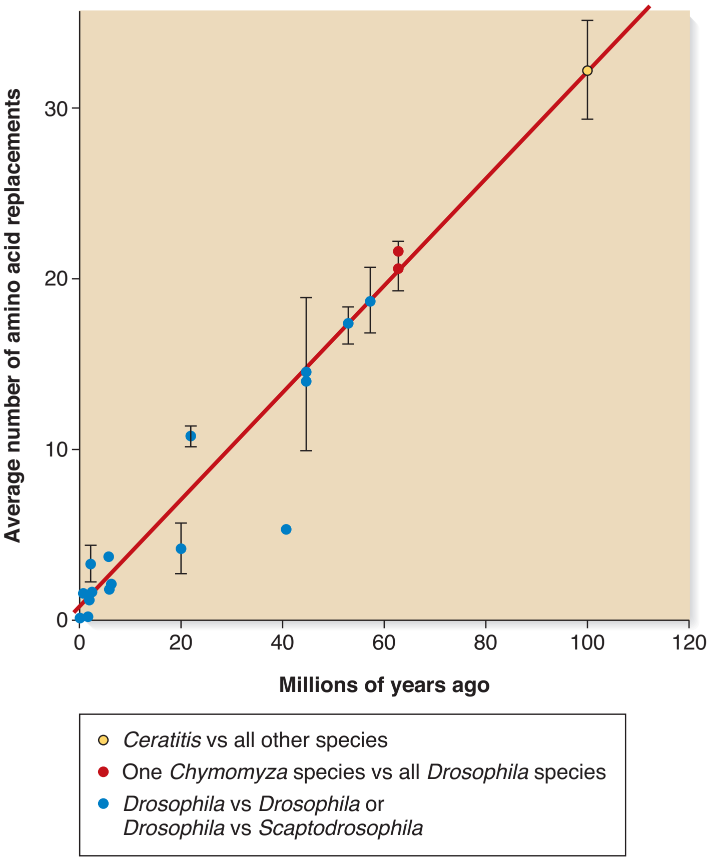
In general, the DNA sequencing of hundreds of thousands of different genes from thousands of species has provided compelling support of the neutral theory of evolution. For example, when comparing the coding sequences within homologous genes of modern species, researchers have found that nucleotide differences are more likely to be observed in the third base (the wobble base) than in the first or second base within a codon. Mutations in the wobble base are often silent, that is, they do not change the amino acid sequence of the protein.

When it was first proposed, the neutral theory sparked a great debate. Some geneticists, called selectionists, strongly opposed the neutralist theory of evolution. However, the debate largely cooled after Tomoko Ohta incorporated the concept of nearly neutral mutations into the theory. Nearly neutral mutations have a minimal effect on phenotype—they may be slightly beneficial or slightly detrimental. Ohta suggested that the prevalence of such alleles can depend mostly on natural selection or mostly on genetic drift, depending on the population size.

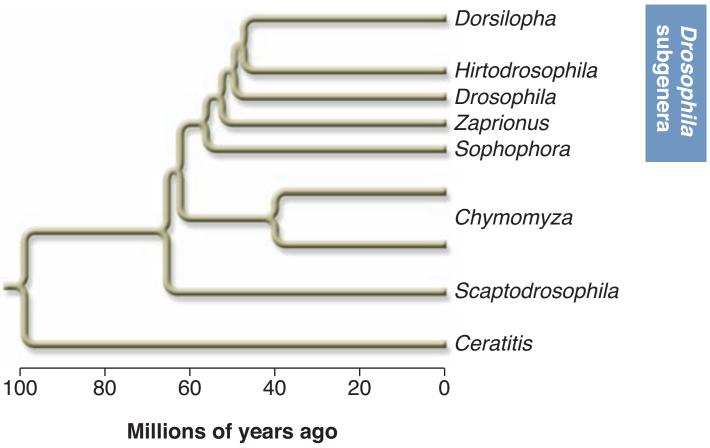
Molecular Clocks Can Be Used to Date the Divergence of Species

Why do evolutionary biologists care about neutral or nearly neutral mutations? One reason is that their prevalence is used as a tool to date evolutionary events. According to the neutral theory of evolution, most of the observed variation is due to neutral mutations. In a sense, the relatively constant rate of neutral or nearly neutral mutations acts as a **molecular clock** on which to measure evolutionary time. According to this idea, neutral mutations become fixed in a population at a rate that is proportional to the rate of mutation per generation. On this basis, the genetic divergence between species that is due to neutral mutations reflects the time elapsed since their last common ancestor.

Figure 27.13a shows an example of a molecular clock involving a study of superoxide dismutase found in 27 different fruit fly species. (Superoxide dismutase protects cells against harmful free radicals.) Twenty-three species were in the genus *Drosophila*, two in the genus *Chymomyza*, one in the genus *Scaptodrosophila*, and one in the genus *Ceratitis*. The genus *Ceratitis* is in the family Tephritidae and is more distantly related to the



(a) An example of a molecular clock



(b) A phylogenetic tree with a timescale

FIGURE 27.13 **A molecular clock.** According to the concept of a molecular clock, neutral or nearly neutral mutations accumulate over evolutionary time at a fairly constant rate. (**a**) The clock shown here is based on an analysis of superoxide dismutase found in 27 different fruit fly species. Twenty-three species were in the genus *Drosophila*, two in the genus *Chymomyza*, one in the genus *Scaptodrosophila*, and one in the genus *Ceratitis*. When comparing the amino acid sequence of this enzyme between species or groups of species, those that diverged more recently tend to have fewer differences than those whose common ancestor occurred in the very distant past. (**b**) A phylogenetic tree for the species shown in part (a). Parts (a) and (b) are modified from F.J. Ayala (1997) Vagaries of the molecular clock. *Proc Natl Acad Sci USA 94*, 7776–7783.

Concept check: With regard to phylogenetic trees, how is a molecular clock useful?

other 26 species, which are in the family Drosophilidae. In this figure, the y-axis is a measure of the average number of amino acid differences in superoxide dismutase between pairs of species or between groups of species. The x-axis plots the amount of time that has elapsed since a pair or two different groups shared a common ancestor. For example, the yellow dot represents the average number of amino acid differences between *Ceratitis* (in family Tephritidae) versus the other species (in family Drosophilidae). Approximately 30 amino acid differences were observed. By comparison, each of the two red dots compares one of the *Chymomyza* species with the *Drosophila* species. The blue dots are pairwise comparisons of *Drosophila* species with each other or with the *Scaptodrosophila* species. As a general trend, species that diverged a longer time ago show a greater number of amino acid differences than species that diverged more recently. The explanation for this phenomenon is that species accumulate independent mutations (e.g., nearly neutral mutations) after they diverge from each other. A longer period of time since their divergence allows for a greater accumulation of mutations that makes their sequences different.

To further understand the concept of a molecular clock, let's consider how the molecular clock data are related to a phylogenetic tree (**Figure 27.13b**). In this diagram, the *Drosophila* genus is divided into five subgenera. The divergence between the genus *Ceratitis* and the other genera occurred a long time ago, nearly 100 mya. The molecular clock data described in Figure 27.13a showed a relatively large number of amino acid differences (about 30) between *Ceratitis* and the other genera. By comparison, the divergence between *Chymomyza* species and *Drosophila* species occurred more recently—about 65 mya as shown in Figure 27.13b. Likewise the molecular clock data revealed fewer amino acid differences of about 20 between *Chymomyza* and *Drosophila* species (see Figure 27.13a). Finally, the five subgenera of *Drosophila* diverged the most recently and had the fewest number of amino acids differences.

Figure 27.13a suggests a linear relationship between the number of sequence changes and the time of divergence. Such a relationship indicates that the observed rate of neutral or nearly neutral mutations remains constant over millions of years. For example, a linear relationship predicts that a pair of species showing 20 nucleotide differences in a given genetic sequence would have a (most recent) common ancestor that existed twice as long ago as another pair showing 10 nucleotide differences. Although actual data, such as the data shown in Figure 27.13a, sometimes show a relatively linear relationship over a defined period, evolutionary biologists have discovered that molecular clocks are often not linear. This is particularly true over very long periods or when comparing distantly related taxa. When comparing different species or groups of species, several factors can contribute to the nonlinearity of molecular clocks. These include the following:

1. Differences in population sizes that may affect the relative effects of genetic drift and natural selection
2. Differences in mutation rates among different species and taxa
3. Differences in the generation times of the species being analyzed
4. Differences in the relative number of sites in a gene or protein that are susceptible to neutral mutations

To produce a timescale, researchers need to calibrate their molecular clocks. How much time does it take to accumulate a certain percentage of nucleotide changes? To perform such a calibration, researchers must have information regarding the date when two species shared a common ancestor. Such information could come from the fossil record, for instance. The genetic differences between those species are then divided by the time elapsed since their last common ancestor to calculate a rate of change. For example, fossil evidence suggests that humans and chimpanzees diverged from a common ancestor approximately 6 mya. The percentage of nucleotide differences in mitochondrial DNA between humans and chimpanzees is 12%. From these data, the molecular clock for changes in mitochondrial DNA sequences of primates is calibrated at roughly 2% nucleotide changes per million years.

Evolution Is Associated with Changes in Chromosome Structure and Number

In this section, we focused on mutations that alter the DNA sequences within genes. In addition to gene mutations, however, other types of changes, such as gene duplications, transpositions, inversions, translocations, and changes in chromosome number, are important features of evolution.

As discussed earlier, changes in chromosome structure and/or number can lead to reproductive isolation and the origin of new species. As an example of variation in chromosome structure among closely related species, **Figure 27.14** compares the banding pattern of the three largest chromosomes in humans and the corresponding chromosomes in chimpanzees, gorillas, and orangutans. The banding patterns are strikingly similar because these species are closely related evolutionarily. However, some interesting differences are observed. Humans have one large chromosome 2, but this chromosome is divided into two separate chromosomes in the other three species. This explains why humans have 23 pairs of chromosomes and the other species have 24. This may have occurred by a fusion of the two smaller chromosomes during the development of the human lineage. Another interesting change in chromosome structure is seen in chromosome 3. The banding patterns among humans, chimpanzees, and gorillas are very similar, but the orangutan has a large inversion that flips the arrangement of bands in the centromeric region.

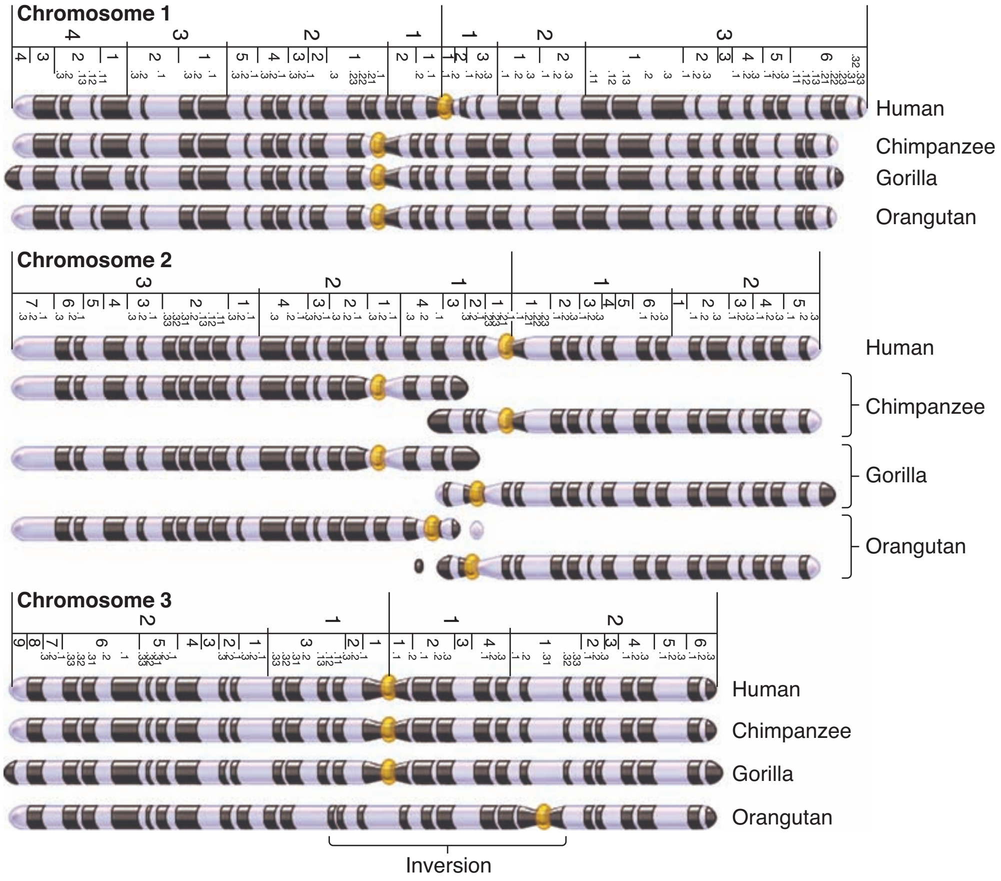


FIGURE 27.14 A comparison of banding patterns among the three largest human chromosomes and the corresponding chromosomes in apes. This is a schematic drawing of late prophase chromosomes. The conventional numbering system of the banding patterns is shown next to the human chromosomes.

Concept check: Describe two differences among these chromosomes.

Synteny Groups Contain the Same Group of Linked Genes

With the advent of molecular techniques, researchers can analyze the chromosomes of two or more different species and identify regions that contain the same groups of linked genes, which are called **synteny groups.** Within a particular synteny group, the same types of genes are found in the same order. In 1995, Graham Moore and colleagues analyzed the locations of molecular markers along the chromosomes of several cereal grasses, including rice (*Oryza sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), foxtail millet (*Setaria italica*), sugarcane (*Saccharum officinarum*), and sorghum (*Sorghum vulgare*). From this analysis, they were able to identify several large synteny groups that are common to most of these species (**Figure 27.15**). As an example, let's compare rice (12 chromosomes per set) with wheat (7 chromosomes per set). In rice, chromosome 6 contains two synteny groups designated R6a and R6b, whereas chromosome 8 contains a single synteny group, R8. In wheat, chromosome 7 consists of R6a and R6b at either end, while R8 is sandwiched in the middle. One possible explanation for this difference could be a chromosomal rearrangement during the evolution of wheat in which R8 became inserted into the middle of a chromosome containing R6a and R6b. Overall, the evolution of cereal grass species has maintained most of the same types of genes, but many chromosomal rearrangements have occurred. These types of rearrangements promote reproductive isolation.

27.3 REVIEWING THE KEY CONCEPTS

- Molecular evolution refers to patterns and processes associated with evolutionary change at the molecular level.
- Homologous genes are derived from the same ancestral gene. They can be orthologs or paralogs (see Figures 27.11, 27.12).
- The neutral theory of evolution suggests that most variation in gene sequences is due to mutations that are neutral or nearly neutral with regard to phenotype. Such variation accumulates in populations largely due to genetic drift.
- The relatively constant rate of neutral or nearly neutral mutations acts as a molecular clock on which to measure evolutionary time (see Figure 27.13).
- Evolution also involves changes in chromosome structure and number (see Figures 27.14, 27.15).

27.3 COMPREHENSION QUESTIONS

1. Homologous genes found in different species are called
   a. orthologs.
   b. paralogs.
   c. analogs.
   d. none of the above.
2. A molecular clock may not be linear because
   a. mutation rates may differ among different species and taxa.
   b. differences in population sizes may affect the relative effects of natural selection and genetic drift.
   c. different species may differ in their generation times.
   d. of all of the above.
3. When comparing the chromosomes of closely related species,
   a. the banding patterns are often similar.
   b. a few structural alterations may be seen.
   c. a change in chromosome number may be seen.
   d. all of the above are common.

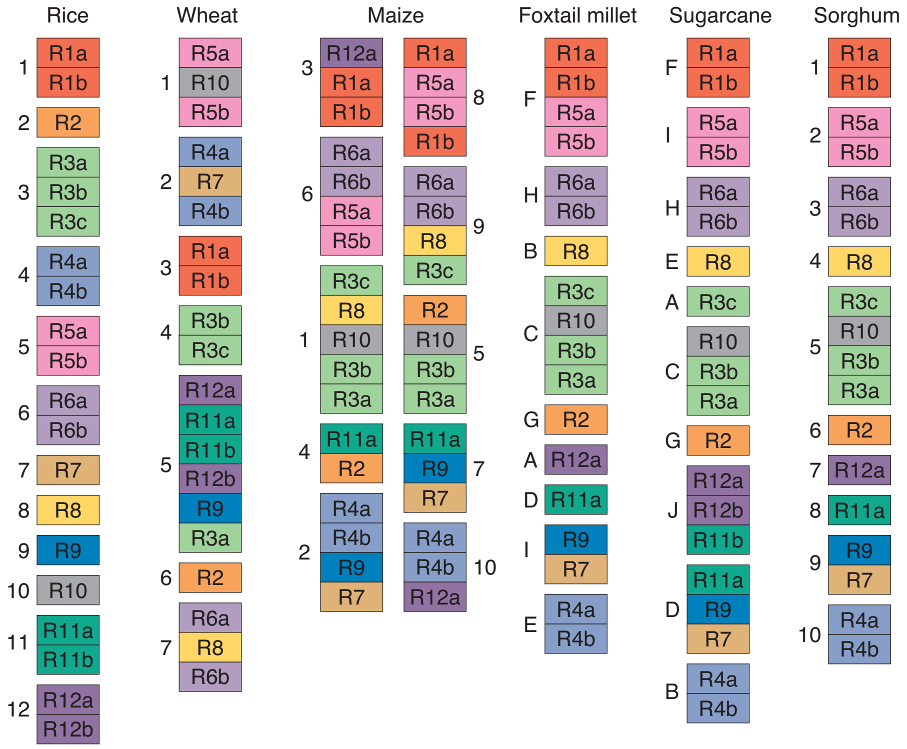

Rice
1 R1a R1b
2 R2
3 R3a R3b R3c
4 R4a R4b
5 R5a R5b
6 R6a R6b
7 R7
8 R8
9 R9
10 R10
11 R11a R11b
12 R12a R12b
Wheat
1 R5a R10 R5b
2 R4a R7 R4b
3 R1a R1b
4 R3b R3c
5 R12a R11a R11b R12b R9 R3a
6 R2
7 R6a R8 R6b
Maize
3 R12a R1a R1b
8 R1a R5a R5b R1b
6 R6a R6b R5a R5b
9 R6a R6b R8 R3c
1 R3c R8 R10 R3b R3a
5 R2 R10 R3b R3a
4 R11a R2
7 R11a R9 R7
2 R4a R4b R9 R7
10 R4a R4b R12a
Foxtail millet
F R1a R1b R5a R5b
H R6a R6b
B R8
C R3c R10 R3b R3a
G R2
A R12a
D R11a
I R9 R7
E R4a R4b
Sugarcane
F R1a R1b
I R5a R5b
H R6a R6b
E R8
A R3c
C R10 R3b R3a
G R2
J R12a R12b R11b
D R11a R9 R7
B R4a R4b
Sorghum
1 R1a R1b
2 R5a R5b
3 R6a R6b
4 R8
5 R3c R10 R3b R3a
6 R2
7 R12a
8 R11a
9 R9 R7
10 R4a R4b

FIGURE 27.15 **Synteny groups in cereal grasses.** The synteny groups are named according to their relative arrangement in rice, which is shown on the left and has 12 chromosomes per set. In some cases, a synteny group or portion of a group may have incurred an inversion, but these are not shown in this figure. Also, the genome of maize contains two copies of most synteny groups, suggesting that it is derived from an ancestral polyploid species.

27.4 EVO-DEVO: EVOLUTIONARY DEVELOPMENTAL BIOLOGY

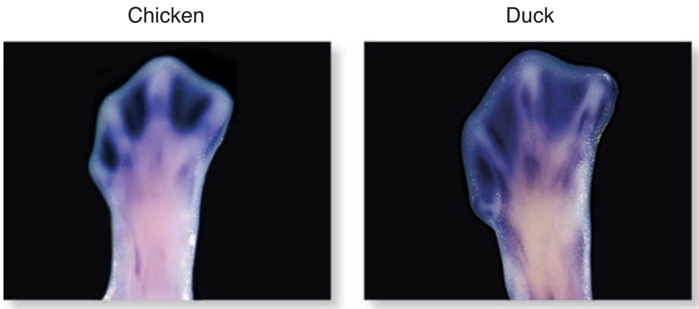
The genetic changes that result in morphological and physiological differences distinguish one species from another. In recent years, researchers have begun to investigate how genetic variation produces species and groups of species with novel shapes and forms. The underlying basis for these changes is often rooted in the developmental pathways that control an organism's morphology. **Evolutionary developmental biology** (informally referred to as **evo-devo**) is a relatively new field of biology that compares the development of species in an attempt to understand ancestral relationships between different species and the developmental mechanisms that bring about evolutionary change. During the past few decades, developmental geneticists have gained a better understanding of biological development at the molecular level. Much of this work has involved the discovery of genes that control development in experimental organisms. As more and more organisms have been analyzed, researchers have become interested in the similarities and differences that occur between closely related and distantly related species. Evolutionary developmental biology has arisen in response to this trend.

How do new morphological forms come into being? For example, how does a nonwebbed foot evolve into a webbed foot? Or how does a new organ, such as an eye, come into existence? Such novelty arises through genetic changes, also called genetic innovations. Certain types of genetic innovation have been so advantageous they have resulted in groups of many new species. For example, the innovation of feathered wings resulted in the evolution of many different species of birds. In this section, we will explore how proteins that control developmental changes, such as cell-signaling proteins and transcription factors, often play a key role in promoting the morphological changes that occur during evolution.

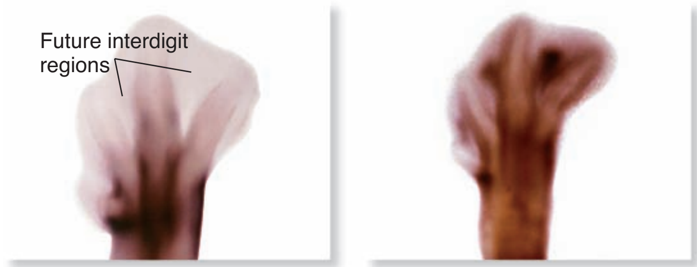
Variation in the Expression Patterns of Developmental Genes Can Dramatically Affect Morphology

As discussed in Chapter 24, genes that play a role in development may influence cell division, cell differentiation, cell migration, and cell death. These four processes produce an organism with a specific body form and function. Developmental genes affect traits such as the size of a flower or the shape of your nose. In recent years, researchers have determined that developmental genes are key players in the evolution of many types of traits. Changes in such genes affect traits that can be acted on by natural selection. Variation in the expression of development genes may be commonly involved in the acquisition of new traits that promote the formation of new species.

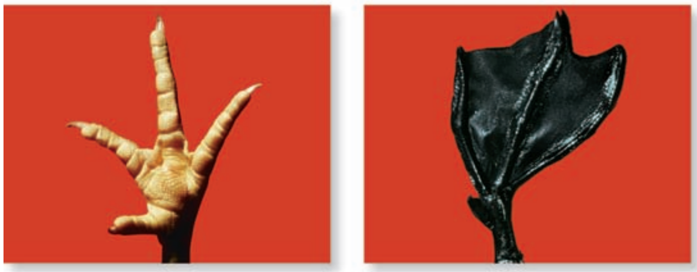
To appreciate this concept, let's compare the formation of a chicken's foot with that of a duck. Developmental biologists have discovered that the morphological differences between a webbed and a nonwebbed foot are due to the differential expression of two cell-signaling proteins called bone morphogenetic protein 4 (BMP4) and gremlin. As shown in **Figure 27.16a**, the *BMP4* gene is expressed throughout the developing limb of both chickens and ducks. The BMP4 protein, which is stained in blue, causes cells to undergo apoptosis and die. The gremlin protein, which is stained brown in **Figure 27.16b**, inhibits the function of BMP4 and thereby allows cells to survive. In the developing chicken limb, the gremlin gene is expressed throughout the limb, except in the regions between each digit. Therefore, these cells die, and a chicken develops a nonwebbed foot (**Figure 27.16c**). In contrast, *gremlin* expression in the duck occurs throughout the entire limb, including the interdigit regions, resulting in a



(a) BMP4 protein levels



(b) Gremlin protein levels



(c) Comparison of a chicken foot and a duck foot

FIGURE 27.16 The role of cell-signaling proteins in the morphology of birds' feet. **(a)** Expression of the *BMP4* gene in the developing limbs. BMP4 protein is stained blue. **(b)** Expression of the *gremlin* gene in the developing limbs. Gremlin protein is stained brown and is expressed in the interdigit region only in the duck. BMP4 causes programmed cell death; Gremlin inhibits BMP4. **(c)** Because BMP4 is not inhibited in the interdigit regions in the chicken, the cells in this region die, and the foot is not webbed. By comparison, inhibition of BMP4 in the interdigit regions in the duck results in a webbed foot.

Concept check: What is the result when Gremlin protein does not inhibit BMP4?

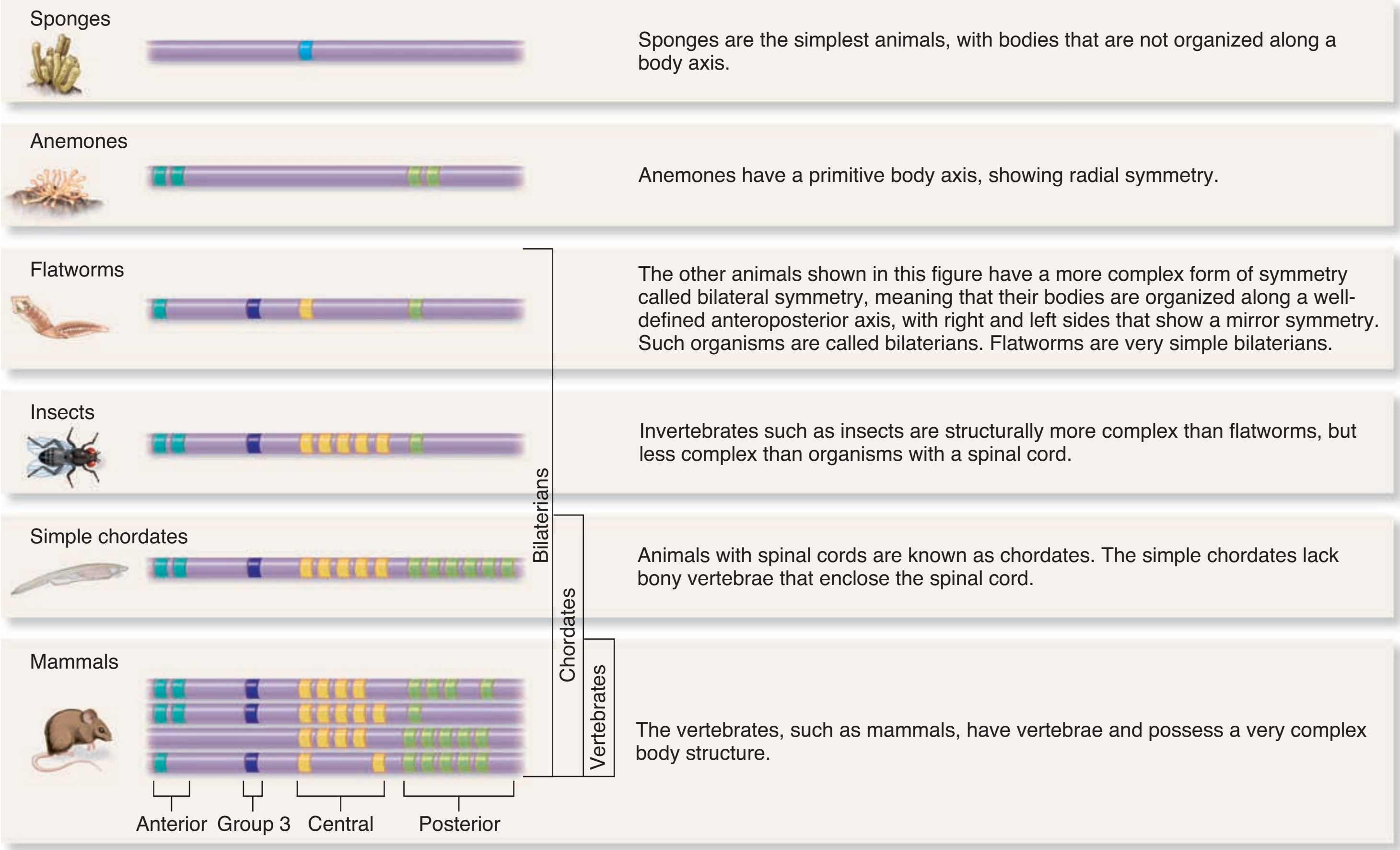
webbed foot. Interestingly, researchers have been able to introduce gremlin protein into the interdigit regions of developing chicken limbs. This produces a chicken with webbed feet!

How do these observations relate to bird evolution? As we have seen, variation in the expression of these genes determines whether or not a bird's feet are webbed. At some point in the evolution of birds, mutations occurred that provided variation in the expression of the *BMP4* and *gremlin* genes, which resulted in nonwebbed or webbed feet. In terrestrial settings, having nonwebbed feet is an advantage because these are more effective at holding on to perches, running along the ground, and snatching prey. Therefore, natural selection maintains nonwebbed feet in terrestrial environments. This process explains the occurrence of nonwebbed feet in chickens, hawks, crows, and many other terrestrial birds. In aquatic environments, webbed feet are an advantage because they act as paddles for swimming, so genetic variation that produced webbed feet has been promoted by natural selection. Over the course of many generations, this gave rise to webbed feet that are now found in ducks, geese, penguins, and other aquatic birds.

How does having webbed or nonwebbed feet influence speciation? This trait may not directly affect the ability of two individuals to mate. However, due to natural selection, birds with webbed feet were more fit and therefore became more prevalent in aquatic environments, whereas birds with nonwebbed feet are common in terrestrial locations. Therefore, reproductive isolation occurs because the populations occupy different environments.

The Evolution of Animal Body Plans Is Related to Changes in *Hox* Gene Number and Expression

Hox genes, which are discussed in Chapter 24, are found in all animals. Developmental biologists have speculated that genetic variation in the *Hox* genes has spawned the formation of many new body types, yielding many different animal species. As shown in **Figure 27.17**, the number and arrangement of *Hox* genes varies considerably among different types of animals. Sponges, the simplest of animals, have at least one *Hox* gene, whereas insects typically have nine or more. In most cases, multiple *Hox* genes occur in a cluster in which the genes are



Sponges are the simplest animals, with bodies that are not organized along a body axis.

Anemones have a primitive body axis, showing radial symmetry.

The other animals shown in this figure have a more complex form of symmetry called bilateral symmetry, meaning that their bodies are organized along a well-defined anteroposterior axis, with right and left sides that show a mirror symmetry. Such organisms are called bilaterians. Flatworms are very simple bilaterians.

Invertebrates such as insects are structurally more complex than flatworms, but less complex than organisms with a spinal cord.

Animals with spinal cords are known as chordates. The simple chordates lack bony vertebrae that enclose the spinal cord.

The vertebrates, such as mammals, have vertebrae and possess a very complex body structure.

FIGURE 27.17 ***Hox* gene composition in different types of animals.** Researchers speculate that the duplication of *Hox* genes and *Hox* gene clusters played a key role in the evolution of more complex body plans in animals. The *Hox* genes are divided into four groups, called anterior, group 3, central, and posterior, based on their relative similarities. Each group is represented by a different color in this figure.

Concept check: With regard to Hox genes and body complexity, what general trend is observed in this figure?

close to each other along a chromosome. In mammals, such *Hox* gene clusters have been duplicated twice during the course of evolution to form four clusters for a total of 38 genes.

How would an increase in *Hox* genes enable more complex body forms to evolve? Part of the answer lies in the spatial expression of the *Hox* genes. In fruit flies, for example, different *Hox* genes are expressed in different segments of the body along the anteroposterior axis (see Figure 24.17). Therefore, an increase in the number of *Hox* genes allows each of these master control genes to become more specialized in the region it controls. One segment in the middle of the fruit fly body can be controlled by a particular *Hox* gene and form wings and legs, whereas a segment in the head region can be controlled by a different *Hox* gene and develop antennae. Therefore, research suggests that one way that new, more complex body forms evolved was through an increase in the number of *Hox* genes, thereby making it possible to form many specialized parts of the body that are organized along a body axis.

Three lines of evidence support the idea that *Hox* gene complexity has been instrumental in the evolution and speciation of animals with different body patterns. First, *Hox* genes are known to control body development. Second, as described in Figure 27.17, a general trend is observed in which animals with a more complex body structure tend to have more *Hox* genes and more *Hox* clusters in their genomes than do simpler animals. Finally, a comparison of *Hox* gene evolution and animal evolution bears striking parallels. Researchers can analyze *Hox* gene sequences among modern species and estimate the timing of past events via molecular clock data. Geneticists can estimate when the first *Hox* gene arose by gene innovation; the date is difficult to pinpoint but is well over 600 mya. The single *Hox* gene found in the sponge descended from this primordial *Hox* gene. In addition, gene duplications of this primordial gene produced clusters of *Hox* genes in other species. Clusters, such as those found in modern insects, were likely to have arisen approximately 600 mya. A duplication of that cluster is estimated to have occurred around 520 mya. Remarkably, these estimates of *Hox* gene origins correlate with major speciation events in the history of animals. The Cambrian explosion, which occurred from 533 to 525 mya, saw a phenomenal diversification in the body plan of invertebrate species. This diversification occurred after the *Hox* cluster was formed and was possibly undergoing its first duplication that resulted in two *Hox* clusters. Also, approximately 420 mya, a second duplication produced species with four *Hox* clusters. This event precedes the proliferation of tetrapods—vertebrates with four limbs—that occurred approximately 417 to 354 mya. Modern tetrapods, such as mammals, have four *Hox* clusters. This second duplication may have been a critical event that led to the evolution of complex terrestrial vertebrates with four limbs, such as lizards, bears, and humans.

The Study of the *Pax6* Gene Indicates That Different Types of Eyes May Have Evolved from a Simpler Form

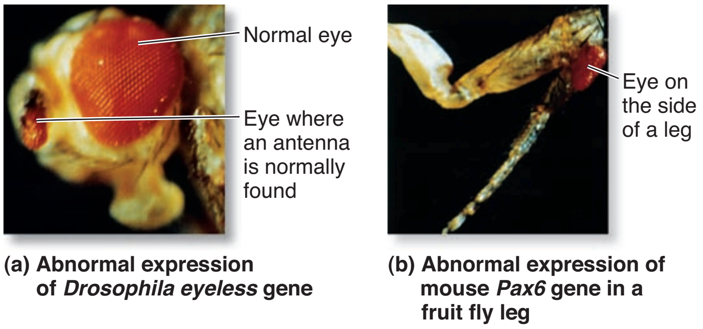
Explaining how a complex organ comes into existence is another major challenge for evolutionary biologists. Although it is relatively easy to understand how a limb could undergo evolutionary modifications to become a wing, flipper, or arm, it is more difficult to understand how a body structure comes into being in the first place. In his book *The Origin of Species*, Charles Darwin addressed this question and pointed out that the existence of complex organs, such as the eye, was difficult to understand. As Darwin noted, the eye of vertebrate species is exceedingly complex, being able to adjust focus, let in different amounts of light, and detect a spectrum of colors. He suggested that such a complex eye must have evolved from a simpler structure through the process of descent with modification. With amazing insight, he speculated that a very simple eye could be composed of just two cells, a photoreceptor cell and an adjacent pigment cell. The photoreceptor cell, which is a type of nerve cell, is able to absorb light and respond to it. The function of the pigment cell is to stop the light from reaching one side of the photoreceptor cell. This primitive, two-cell arrangement would allow an organism to sense both light and the direction from which the light comes.

How would natural selection play a role in the evolution of eyes? A primitive eye would provide an additional way for a mobile organism to sense its environment, possibly allowing it to avoid predators or locate food. Vision is nearly universal among animals, which indicates there must be a strong selective advantage to better eyesight. Over time, eyes could become more complex by enhancing the ability to absorb different amounts and wavelengths of light and also by refinements in structures such as the addition of lenses that detect the direction of light.

Since the time of Darwin, many evolutionary biologists have wrestled with the question of eye evolution. From an anatomical point of view, researchers have discovered many different types of eyes. For example, the eyes of fruit flies, squid, and humans are quite different from each other. Furthermore, species that are closely related evolutionarily sometimes have different types of eyes. This observation led evolutionary biologists Luitfried von Salvini-Plawen and Ernst Mayr to propose that eyes may have independently arisen many different times during evolution. Based solely on morphology, such a hypothesis seemed reasonable, and for many years was accepted by the scientific community.

The situation took a dramatic turn when geneticists began to study eye development. Researchers identified a gene, *Pax6,* that influences eye development in both rodents and humans. In mice and rats, a mutation in this gene results in small eyes. In humans, a mutation in the *Pax6* gene results in an eye disorder called aniridia. In heterozygotes carrying one defective copy of the gene, the iris does not form. Researchers subsequently discovered a gene in *Drosophila*, named *eyeless*, that also causes a defect in eye development when mutant. DNA sequencing revealed that the *eyeless* and *Pax6* genes are homologous; they are derived from the same ancestral gene.

In 1995, Walter Gehring and his colleagues were able to show experimentally that the abnormal expression of the *eyeless* gene in other parts of the fruit fly body could promote the formation of additional eyes. For example, using genetic engineering techniques, they were able to express the *eyeless* gene in the region where antennae should form. As seen in **Figure 27.18a**, this resulted in the formation of fruit fly eyes where antennae


Normal eye
Eye where an antenna is normally found
Eye on the side of a leg

(a) Abnormal expression of *Drosophila eyeless* gene **(b) Abnormal expression of mouse *Pax6* gene in a fruit fly leg**

FIGURE 27.18 **Formation of additional eyes in *Drosophila* due to the abnormal expression of a master control gene for eye morphogenesis.** **(a)** When the *Drosophila eyeless* gene is expressed in the antenna region, eyes are formed where antennae should be located. **(b)** When the mouse *Pax6* gene is expressed in the leg region of *Drosophila*, a small eye is formed there. Note: *Pax* is an acronym for paired box. The protein encoded by this gene contains a domain called a paired box.

Concept check: In part b, why is a Drosophila eye produced in an odd place rather than a mouse eye?

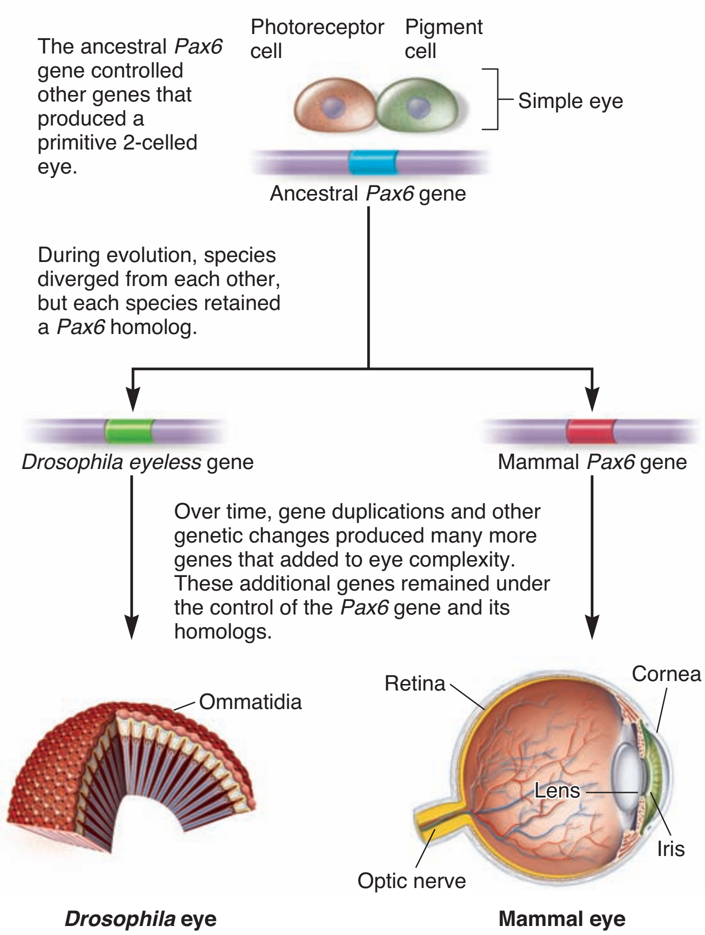

The ancestral *Pax6* gene controlled other genes that produced a primitive 2-celled eye.
Photoreceptor cell
Pigment cell
Simple eye
Ancestral *Pax6* gene
During evolution, species diverged from each other, but each species retained a *Pax6* homolog.
Drosophila eyeless gene
Mammal *Pax6* gene
Over time, gene duplications and other genetic changes produced many more genes that added to eye complexity. These additional genes remained under the control of the *Pax6* gene and its homologs.
Ommatidia
Retina
Cornea
Lens
Iris
Optic nerve
Drosophila eye
Mammal eye

FIGURE 27.19 **Eye evolution.** In this diagram, genetic changes, under the control of the ancestral *Pax6* gene, led to the evolution of different types of eyes. Ommatidia are units of the insect compound eye.

are normally found! Remarkably, the expression of the mouse *Pax6* gene in *Drosophila* can also cause the formation of eyes in unusual places. **Figure 27.18b** shows the formation of an eye on the leg of a fruit fly. This eye was caused by the expression of the mouse *Pax6* gene in this region.

The mouse *Pax6* gene switches on eye formation in *Drosophila*, but the eye produced is a *Drosophila* eye, not a mouse eye. This happens because the genes activated by the *Pax6* gene are all from the *Drosophila* genome. In *Drosophila*, the *eyeless* gene switches on a cascade involving 2500 different genes required for eye morphogenesis. The *Pax6* gene and its *Drosophila* homolog, *eyeless,* are master control genes that promote the formation of an eye. Based on these observations, Gehring has suggested that the eyes of *Drosophila* and mammals are evolutionarily derived from the modification of an eye that arose once during evolution. If *Drosophila* and mammalian eyes had arisen independently, the *Pax6* gene from mice would not be expected to induce the formation of eyes in *Drosophila*. Alternatively, other researchers argue that *Pax6* may control only certain features of eye development and that different types of eyes may have evolved independently. Future research will be needed to resolve this controversy.

Since the initial discovery of the *Pax6* and *eyeless* genes, homologs of this gene have been discovered in many different species. In all cases where it has been tested, this gene directs eye development. The *Pax6* gene and its homologs encode a transcription factor protein that controls the expression of many different genes. Gehring and colleagues have hypothesized that the eyes from many different species all evolved from a common ancestral form consisting of, as proposed by Darwin, one photoreceptor cell and one pigment cell (**Figure 27.19**). As mentioned previously, such a very simple eye can accomplish some rudimentary form of vision by detecting light and its direction. Eyes such as these are still found in modern species such as the larvae of certain types of mollusks. Over the course of evolution, simple eyes were transformed into more complex types of eyes by modifications that resulted in the addition of more types of cells such as lens and muscle cells.

27.4 REVIEWING THE KEY CONCEPTS

- Evolutionary developmental biology is a relatively new field of biology that compares the development of species in an attempt to understand ancestral relationships between different species and the developmental mechanisms that bring about evolutionary change.
- Variation in the pattern of expression of developmentally important genes, such as those involved with cell signaling, can have a dramatic effect on morphology (see Figure 27.16).
- The evolution of animal body plans has involved changes in the number and expression patterns of *Hox* genes (see Figure 27.17).
- The *Pax6* gene and its *Drosophila* homolog, *eyeless,* are master control genes that promote the formation of an eye. This observation suggests that the eyes of all animals may be evolutionarily derived from the modification of an eye that arose once during evolution (see Figures 27.18, 27.19).

27.4 COMPREHENSION QUESTIONS

1. Proteins that commonly are involved in promoting developmental changes include
   a. transcription factors.
   b. cell-signaling proteins.
   c. ribosomal proteins.
   d. both a and b.
2. Evidence that *Hox* genes have been instrumental in the evolution and speciation of animals includes which of the following?
   a. *Hox* genes are known to control development along the anteroposterior axis.
   b. Animals with more complex body plans tend to have more *Hox* genes.
   c. Changes in *Hox* gene number are associated with past speciation events.
   d. All of the above are evidence.
3. The *Pax6* gene encodes a _____________ that controls ____________.
   a. transcription factor, eye development
   b. cell-signaling protein, eye development
   c. transcription factor, the anteroposterior axis
   d. cell-signaling protein, the anteroposterior axis

KEY TERMS

Page 672. biological evolution, evolution, microevolution, macroevolution, molecular evolution
Page 673. adaptive evolution, species
Page 674. subspecies, ecotype
Page 675. reproductive isolation, prezygotic isolating mechanisms, postzygotic isolating mechanisms
Page 676. species concept, biological species concept, evolutionary lineage concept, ecological species concept, general lineage concept, speciation, anagenesis, cladogenesis, allopatric speciation
Page 677. founder effect
Page 678. parapatric speciation, hybrid zones, sympatric speciation
Page 679. phylogeny, phylogenetic tree, systematists, homology
Page 680. homologous, monophyletic group, clade
Page 681. phenetic approach, phenograms, cladistic approach, cladograms, ancestral characters, primitive characters, shared derived characters, synaptomorphy, ingroup, outgroup
Page 682. principle of parsimony
Page 684. maximum likelihood, Bayesian methods, Bacteria, Archaea, Eukaryotes, supergroups, vertical evolution, horizontal gene transfer
Page 686. molecular evolution, orthologs, paralogs, gene family
Page 688. nonneutral mutation, neutral theory of evolution, molecular clock
Page 691. synteny groups
Page 692. evolutionary developmental biology (evo-devo)

CHAPTER SUMMARY

- Evolution is a heritable change in one or more characteristics of a population or species from one generation to the next.

27.1 Origin of Species

- The most commonly used characteristics for distinguishing species are morphological traits, the ability to interbreed, molecular features, ecological factors, and evolutionary relationships. However, each approach has its drawbacks (see Figure 27.1, Table 27.1).
- A species concept is a way of defining the concept of a species and/or providing an approach for distinguishing one species from another. The general lineage concept is receiving wide support.
- Speciation is the process by which new species are formed via evolution. It may occur by anagenesis, when one species evolves into a new one, or more commonly by cladogenesis, in which a species diverges into two or more different species (see Figure 27.2).
- Cladogenesis may occur via allopatric, parapatric, or sympatric speciation (see Table 27.2, Figures 27.3, 27.4).

27.2 Phylogenetic Trees

- A phylogenetic tree is a diagram that describes a phylogeny—the sequence of events involved in the evolutionary development of a species or group of species (see Figure 27.5).
- Different methods, including cladistics and phenetics, are used to construct phylogenetic trees. One way to evaluate the validity of possible trees is by the principle of parsimony (see Figures 27.6, 27.7, 27.8).
- Maximum likelihood and Bayesian methods are also used to discriminate among possible phylogenetic trees.
- Phylogenetic trees help us to understand the relationships among different taxa such as domains and supergroups (see Figure 27.9).
- In addition to vertical evolution, horizontal gene transfer has played an important role in the evolution of life on Earth (see Figure 27.10).

27.3 Molecular Evolution and Molecular Clocks

- Molecular evolution refers to patterns and processes associated with evolutionary change at the molecular level.

- Homologous genes are derived from the same ancestral gene. They can be orthologs or paralogs (see Figures 27.11, 27.12).
- The neutral theory of evolution suggests that most variation in gene sequences is due to mutations that are neutral or nearly neutral with regard to phenotype. Such variation accumulates in populations largely due to genetic drift.
- The relatively constant rate of neutral or nearly neutral mutations acts as a molecular clock on which to measure evolutionary time (see Figure 27.13).
- Evolution also involves changes in chromosome structure and number (see Figures 27.14, 27.15).

27.4 Evo-Devo: Evolutionary Developmental Biology

- Evolutionary developmental biology is a relatively new field of biology that compares the development of species in an attempt to understand ancestral relationships between different species and the developmental mechanisms that bring about evolutionary change.
- Variation in the pattern of expression of developmentally important genes, such as those involved with cell signaling, can have a dramatic effect on morphology (see Figure 27.16).
- The evolution of animal body plans has involved changes in the number and expression patterns of *Hox* genes (see Figure 27.17).
- The *Pax6* gene and its *Drosophila* homolog, *eyeless,* are master control genes that promote the formation of an eye. This observation suggests that the eyes of all animals may be evolutionarily derived from the modification of an eye that arose once during evolution (see Figures 27.18, 27.19).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. A codon for leucine is UUA. A mutation causing a single-base substitution in a gene can change this codon in the transcribed mRNA into GUA (valine), AUA (isoleucine), CUA (leucine), UGA (stop), UAA (stop), UCA (serine), UUG (leucine), UUC (phenylalanine), or UUU (phenylalanine). According to the neutral theory, which of these mutations would you expect to see within the genetic variation of a natural population? Explain.

Answer: The neutral theory proposes that neutral mutations accumulate to the greatest extent in a population. Leucine is a nonpolar amino acid. For a UUA codon, single-base changes of CUA and UUG are silent, so they would be the most likely to occur in a natural population. Likewise, conservative substitutions to other nonpolar amino acids such as isoleucine (AUA), valine (GUA), and phenylalanine (UUC and UUU) may not affect protein structure and function, so they may also occur and not be eliminated rapidly by natural selection. The polar amino acid serine (UCA) is a nonconservative substitution; one would predict that it is more likely to disrupt protein function. Therefore, it may be less likely to be found. Finally, the stop codons, UGA and UAA, would be expected to diminish or eliminate protein function, particularly if they occur early in the coding sequence. These types of mutations are selected against and, therefore, are not usually found in natural populations.

S2. Explain why homologous genes have sequences that are similar but not identical.

Answer: Homologous genes are derived from the same ancestral gene. Therefore, as a starting point, they had identical sequences. Over time, however, each gene accumulates random mutations that the other homologous genes did not acquire. These random mutations change the gene from its original sequence. Therefore, much of the sequence between homologous genes remains identical, but some of the sequence is altered due to the accumulation of independent random mutations.

S3. Explain why plants are more likely to evolve by sympatric speciation compared to animals.

Answer: A common way for sympatric speciation to occur is by the formation of polyploids. For example, if one species is diploid ($2n$), nondisjunction could produce an individual that is tetraploid ($4n$). If the tetraploid individual was a plant and if it was monoecious (i.e., produces both pollen and egg cells), the plant could multiply to produce many tetraploid offspring. These offspring would be reproductively isolated from the diploid plants that are found in the same geographic area. This isolation occurs because the offspring of a cross between a diploid and tetraploid plant would be infertile. For example, if the pollen from a diploid plant fertilized the egg from a tetraploid plant, the offspring would be triploid. The triploid offspring might be viable, but it is likely that triploids would be sterile because they would produce highly aneuploid gametes. The gametes would be aneuploid because there are an odd number of homologous chromosomes. In this case, there would be three copies of each homologous chromosome, and these could not be equally distributed into gametes. Therefore, when a tetraploid is produced, it can immediately become its own unique species that is reproductively isolated from other species.

In contrast, polyploidy rarely occurs in animals. Perhaps the main reason is because animals usually cannot tolerate polyploidy. Tetraploid animals typically die during early stages of development. In addition, most species of animals have male and female sexes. To develop a tetraploid species from a diploid species, nondisjunction would have to produce both a male and female offspring that could reproduce with each other. The chance of producing one tetraploid offspring is relatively rare. The chances of producing two tetraploid offspring that happen to be male and female and happen to mate with each other to produce many offspring would be extremely rare.

S4. As described in Figure 27.14, evolution is associated with changes in chromosome structure and number. As seen there, chromosome 2 in humans is divided into two distinct chromosomes in chimpanzees, gorillas, and orangutans. In addition, chromosome 3 in the orangutan has a large inversion not found in the other three primates. Discuss the potential role of these types of changes in the evolution of these primate species.

Answer: As discussed in Chapter 8, changes in chromosome structure, such as inversions and balanced translocations, may not have any phenotypic effects. Likewise, the division of a single chromosome into two distinct chromosomes may not have any phenotypic effect as long as the total amount of genetic material remains the same. Overall, the types of changes in chromosome structure and number shown in Figure 27.14 may not have caused any changes in the phenotypes of primates. However, the changes would be expected to promote reproductive isolation. For example, if a gorilla mated with an orangutan, the offspring would be an inversion heterozygote for chromosome 3. As shown in Figure 8.10, crossing over during gamete formation in an inversion heterozygote may produce chromosomes that have too much or too little genetic material. This is particularly likely if the inversion is fairly large (e.g., the one shown in Figure 27.14). The inheritance of too much or too little genetic material is likely to be detrimental or even lethal. For this reason, the hybrid offspring of a gorilla and orangutan would probably not be fertile. (Note: In reality, there are several other reasons why interspecies matings between gorillas and orangutans do not produce viable offspring.)

Overall, the primary effect of changes in chromosome structure and number, like the ones shown in Figure 27.14, is to promote reproductive isolation. Once two populations become reproductively isolated, they accumulate different mutations, which over the course of many generations leads to two different species with distinct characteristics.

Conceptual Questions

C1. Discuss the two principles on which Darwin's theory of evolution is based.

C2. Evolution, which involves genetic change in a population of organisms over time, is often described as the unifying theme in biology. Discuss how evolution is unifying at the molecular and cellular levels.

C3. What is a species? What types of observations do researchers analyze when trying to identify species?

C4. What is meant by the term *reproductive isolation*? Give several examples.

C5. Would the following examples of reproductive isolation be considered a prezygotic or postzygotic mechanism?

A. Horses and donkeys can interbreed to produce mules, but the mules are infertile.

B. Three species of the orchid genus *Dendrobium* produce flowers 8 days, 9 days, and 11 days after a rainstorm. The flowers remain open for 1 day.

C. Two species of fish release sperm and eggs into seawater at the same time, but the sperm of one species do not fertilize the eggs of the other species.

D. Two tree frogs, *Hyla chrysoscelis* (diploid) and *H. versicolor* (tetraploid), can produce viable offspring, but the offspring are sterile.

C6. Distinguish between anagenesis and cladogenesis. Which type of speciation is more prevalent? Why?

C7. Describe three or more genetic mechanisms that may lead to the rapid evolution of a new species. Which of these genetic mechanisms are the result of natural selection, and which are not?

C8. Explain the type of speciation (allopatric, parapatric, or sympatric) most likely to occur under each of the following conditions:

A. A pregnant female rat is transported by an ocean liner to a new continent.

B. A meadow containing several species of grasses is exposed to a pesticide that promotes nondisjunction.

C. In a very large lake containing several species of fish, the water level gradually falls over the course of several years. Eventually, the large lake becomes subdivided into smaller lakes, some of which are connected by narrow streams.

C9. Alloploids are produced by crosses involving two different species. Explain why alloploids are reproductively isolated from the two original species from which they were derived. Explain why alloploids are usually sterile, whereas allotetraploids (containing a diploid set from each species) are commonly fertile.

C10. Discuss whether the phenomenon of reproductive isolation applies to bacteria, which reproduce asexually. How would a geneticist divide bacteria into separate species?

C11. Discuss the major differences among allopatric, parapatric, and sympatric speciation.

C12. The following are two DNA sequences from homologous genes:

TTGCATAGGCATACCGTATGATATCGAAAACTAGAAAAATAGGGCGATAGCTA

GTATGTTATCGAAAAGTAGCAAAATAGGGCGATAGCTACCCAGACTACCGGAT

The two sequences, however, do not begin and end at the same location. Try to line them up according to their homologous regions.

C13. What is meant by the term *molecular clock*? How is this concept related to the neutral theory of evolution?

C14. Would the rate of deleterious or beneficial mutations be a good molecular clock? Why or why not?

C15. Which would you expect to exhibit a faster rate of evolutionary change: the nucleotide sequence of a gene or the amino acid sequence of the encoded polypeptide of the same gene? Explain your answer.

C16. When comparing the coding region of structural genes among closely related species, it is commonly found that certain regions of the gene have evolved more rapidly (i.e., have tolerated more changes in sequence) than other regions of the gene. Explain why different regions of a structural gene evolve at different rates.

C17. Plant seeds contain storage proteins that are encoded by plant genes. When the seed germinates, these proteins are rapidly hydrolyzed (i.e., the covalent bonds between amino acids within the polypeptides are broken), which releases amino acids for the developing seedling. Would you expect the genes that encode plant storage proteins to evolve slowly or rapidly compared with genes that encode enzymes? Explain your answer.

C18. Based on your understanding of molecular genetics (in this and other chapters), describe three or more observations that suggest that all life forms on Earth evolved from a common ancestor or group of ancestors.

C19. Take a look at the α-globin and β-globin sequences in Figure 27.11. Which sequences are more similar, the α-globin in humans and the α-globin in horses, or the α-globin in humans and the β-globin in humans? Based on your answer, would you conclude that the gene duplication that gave rise to the α-globin and β-globin genes occurred before or after the divergence of humans and horses? Explain your reasoning.

C20. Compare and contrast the neutral theory of evolution versus the Darwinian (i.e., selectionist) theory of evolution. Explain why the neutral theory of evolution is sometimes called non-Darwinian evolution.

C21. For each of the following examples, discuss whether it would be the result of neutral mutations or mutations that have been acted on by natural selection, or both:

A. When comparing sequences of homologous genes, differences in the coding sequence are most common at the wobble base (i.e., the third base in each codon).

B. For a structural gene, the regions that encode portions of the polypeptide that are vital for structure and function are less likely to display mutations than are other regions of the gene.

C. When comparing the sequences of homologous genes, introns usually have more sequence differences than exons do.

C22. As discussed in Chapter 25, genetic variation is prevalent in natural populations. This variation is revealed in the DNA sequencing of genes. According to the neutral theory of evolution, discuss the relative importance of natural selection against detrimental mutations, natural selection in favor of beneficial mutations, and neutral mutations in accounting for the genetic variation we see in natural populations.

C23. If you were comparing the karyotypes of species that are closely related evolutionarily, what types of similarities and differences would you expect to find?

C24. In the developing bud that gives rise to a hand in a human embryo, where would you expect the *gremlin* gene to be expressed?

C25. Discuss how *Hox* gene number is related to body complexity.

Application and Experimental Questions

E1. Two populations of snakes are separated by a river. The snakes cross the river only on rare occasions. The snakes in the two populations look very similar to each other except that the members of the population on the eastern bank of the river have a yellow spot on the top of their head, whereas the members of the western population have an orange spot on the top of their head. Discuss two experimental methods that you might follow to determine whether the two populations are members of the same species or members of different ones.

E2. Sympatric speciation by allotetraploidy has been proposed as a common mechanism for speciation. Let's suppose you were interested in the origin of certain grass species in southern California. Experimentally, how would you go about determining if some of the grass species are the result of allotetraploidy?

E3. Two diploid species of closely related frogs, which we will call species A and species B, were analyzed with regard to genes that encode an enzyme called hexokinase. Species A has two distinct copies of this gene: *A1* and *A2*. In other words, this diploid species is *A1A1 A2A2*. The other species has three copies of the hexokinase gene, which we will call *B1*, *B2*, and *B3*. A diploid individual of species B would be *B1B1 B2B2 B3B3*. These hexokinase genes from the two species were subjected to DNA sequencing, and the percentage of sequence identity was compared among these genes. The results are shown here:

Percentage of DNA Sequence Identity

	A1	***A2***	***B1***	***B2***	***B3***
A1	100	62	54	94	53
A2	62	100	91	49	92
B1	54	91	100	67	90
B2	94	49	67	100	64
B3	53	92	90	64	100

If we assume that hexokinase genes were never lost in the evolution of these frog species, how many distinct hexokinase genes do you think there were in the most recent ancestor that preceded the divergence of these two species? Explain your answer. Also explain why species B has three distinct copies of this gene, whereas species A has only two.

E4. A researcher sequenced a portion of a bacterial gene and obtained the following sequence, beginning with the start codon, which is underlined:

ATG CCG GAT TAC CCG GTC CCA AAC AAA ATG ATC GGC CGC CGA ATC TAT CCC

The bacterial strain that contained this gene has been maintained in the laboratory and grown serially for many generations. Recently, another person working in the laboratory isolated DNA from the bacterial strain and sequenced the same region. The following results were obtained.

ATG CCG GAT TAT CCG GTC CCA AAT AAA ATG ATC GGC CGC CGA ATC TAC CCC

Explain why these sequencing differences may have occurred.

E5. F_1 hybrids between two species of cotton, *Gossypium barbadense* and *G. hirsutum*, are very vigorous plants. However, F_1 crosses produce many seeds that do not germinate and a low percentage of very weak F_2 offspring. Suggest two reasons for these observations.

E6. A species of antelope contains 20 chromosomes per set. The species is divided by a mountain range into two separate populations, which we will call the eastern and western population. When comparing the karyotypes of these two populations, it was discovered that the members of the eastern population are homozygous for a large inversion within chromosome 14. How would this inversion affect the interbreeding between the two populations? Could such an inversion play an important role in speciation?

E7. Explain why molecular techniques were needed as a way to provide evidence for the neutral theory of evolution.

E8. In Chapter 21, we learned about a technique called fluorescence in situ hybridization (FISH), during which a labeled piece of DNA is hybridized to a set of chromosomes. Let's suppose that we cloned a piece of DNA from *G. pubescens* (see Figure 27.4) and used it as a labeled probe for in situ hybridization. What would you expect to happen if we hybridized it to the *G. speciosa*, the natural *G. tetrahit*, or the artificial *G. tetrahit* strains? Describe your expected results.

E9. Discuss how the principle of parsimony can be used in a cladistics approach of constructing a phylogenetic tree.

E10. A homologous DNA region, which was 20,000 bp in length, was sequenced among four different species. The following number of nucleotide differences were obtained:

	Species A	Species B	Species C	Species D
Species A	0	443	719	465
Species B	443	0	744	423
Species C	719	744	0	723
Species D	465	423	723	0

Construct a phylogenetic tree that describes the evolutionary relationships among these four species using the UPGMA method. Your tree should include values that show the percentage of nucleotide substitutions.

E11. As discussed in this chapter and Chapter 25, genes are sometimes transferred between different species via horizontal gene transfer. Discuss how horizontal gene transfer might give misleading results when constructing a phylogenetic tree. How could you overcome this problem?

E12. If a researcher used genetic engineering techniques to express the *Drosophila eyeless* gene in the embryo at the region that will become the tip of the mouse's tail, what results would you expect in the resulting offspring?

Questions for Student Discussion/Collaboration

1. The raw material for evolution is random mutation. Discuss whether or not you view evolution as a random process.
2. Compare the forms of speciation that are slow with those that occur more rapidly. Make a list of the slow and fast forms. With regard to mechanisms of genetic change, what features do slow and rapid speciation have in common? What features are different?
3. Do you think that Darwin would object to the neutral theory of evolution?

Answers to Comprehension Questions

27.1: d, f, a, c

27.2: c, d, d, c

27.3: a, d, d

27.4: d, d, a

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookerconcepts

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

APPENDIX
EXPERIMENTAL TECHNIQUES

A

OUTLINE

A.1 Methods for Growing Cells
A.2 Microscopy
A.3 Separation Methods
A.4 Methods for Measuring Concentrations of Molecules and Detecting Radioisotopes and Antigens

A.1 METHODS FOR GROWING CELLS

Researchers often grow cells in a laboratory as a way to study their properties. This is known as a **cell culture.** Cell culturing offers several technical advantages. The primary advantage is that the growth medium is defined and can be controlled. Minimal growth medium contains the bare essentials for cell growth: salts, a carbon source, an energy source, essential vitamins, amino acids, and trace elements. In their experiments, geneticists often compare strains that can grow in minimal media and mutant strains that cannot grow unless the medium is supplemented with additional components. A rich growth medium contains many more components than are required for growth.

Researchers also add substances to the culture medium for other experimental reasons. For example, radioactive isotopes can be added to the culture medium to radiolabel cellular macromolecules. In addition, an experimenter could add a hormone to the growth medium and then monitor the cells' response to the hormone. In all of these cases, cell culturing is advantageous because the experimenter can control and vary the composition of the growth medium.

The first step in creating a cell culture is the isolation of a cell population that the researchers wish to study. For bacteria, such as *Escherichia coli,* and eukaryotic microorganisms, such as yeast and *Neurospora,* the researchers simply obtain a sample of cells from a colleague or a stock center. For animal or plant tissues, the procedure is a bit more complicated. When cells are contained within a complex tissue, they must first be dispersed by treating the tissue with agents that separate it into individual cells to create a cell suspension.

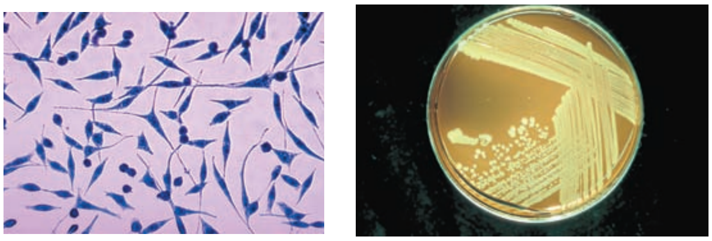
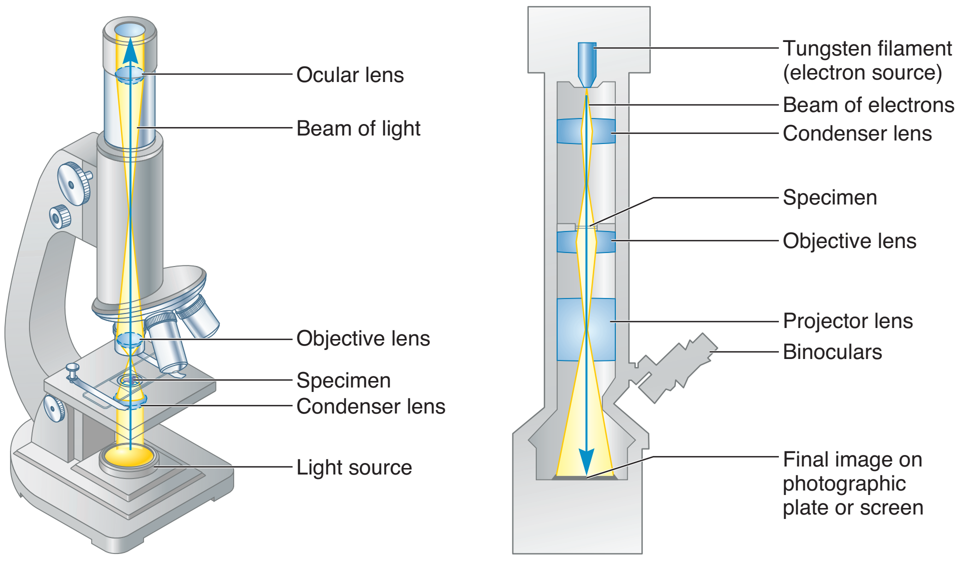


FIGURE A.1 Growth of cells on solid growth media. **(a)** This micrograph shows fibroblasts growing as a monolayer on a solid growth medium. **(b)** Bacterial cells form colonies that are a clonal population of cells derived from a single cell.

Once a desired population of cells has been obtained, researchers can grow them in a laboratory (i.e., in vitro) either suspended in a liquid growth medium or attached to a solid surface such as agar. Both methods have been commonly used in the experiments considered throughout this textbook. Liquid culture is often used when researchers want to obtain a large quantity of cells and isolate individual cellular components, such as nuclei or DNA. By comparison, **Figure A.1** shows animal cells and bacteria cells that are grown on solid growth media. As shown in Figure 22.11, solid media are used to study cancer cells, because such cells can be distinguished by the formation of foci in which malignant cells pile up on top of each other. In gene-cloning experiments with bacteria and yeast, solid media are also used. Each colony of cells is a clone of cells that is derived from a single cell that divided to produce many cells (**Figure A.1b**). As discussed in Chapter 18, a solid medium is used in the isolation of individual clones that contain a desired gene.

A.2 MICROSCOPY

Microscopy is a technique for observing things that cannot be seen (or hardly be seen) with the naked eye. A key concept in microscopy is **resolution,** which is the minimum distance between two objects that enables them to be seen as separate from each other. The ability to resolve two points as being separate depends on several factors, including the wavelength of the illumination source (light or electron beam), the medium in



(a) Light microscope

(b) Transmission electron microscope

FIGURE A.2 **Design of (a) optical (light) and (b) transmission electron microscopes.**

which the sample is immersed, and the structural features of the microscope (which are beyond the scope of this textbook).

As shown in **Figure A.2**, two widely used kinds of microscopes are the optical (light) microscope and the transmission electron microscope (TEM). The light microscope is used to resolve cellular structures to a limit of approximately 0.3 μm. (For comparison, a typical bacterium is about 1 μm long.) At this resolution, the individual cell organelles in eukaryotic cells can be discerned easily, and chromosomes are also visible. Karyotyping is accomplished via light microscopy after the chromosomes have been treated with stains. A variation of light microscopy known as fluorescence microscopy is often used to highlight a particular feature of a chromosome or cellular structure. The technique of fluorescence in situ hybridization (FISH; see Chapter 20) makes use of this type of microscope. Also, optical modifications in certain light microscopes (e.g., phase contrast and differential interference) can be used to exaggerate the differences in densities between neighboring cells or cell structures. These kinds of light microscopes are useful in monitoring cell division in living (unstained) cells or in transparent worms (as in Figure 23.16).

The structural details of large macromolecules such as DNA and ribosomes are not observable by light microscopy. The coarse topology of these macromolecules can be determined by electron microscopy. Electron microscopes have a limit of resolution of about 2 nm, which is about 100 times finer than the best light microscopes. The primary advantage of electron microscopy over light microscopy is its better resolution. Disadvantages include a much higher expense and more extensive sample preparation. In transmission electron microscopy, the sample is bombarded with an electron beam. This requires that the sample be dried, fixed, and usually coated with a heavy metal that absorbs electrons.

A.3 SEPARATION METHODS

Biologists often wish to take complex systems and separate them into less complex components. For example, the cells within a complex tissue can be separated into individual cells, or the macromolecules within cells can be separated from the other cellular components. In this section, we focus primarily on methods aimed at separating and purifying macromolecules.

Disruption of Cellular Components

In many experiments described in this textbook, researchers have obtained a sample of cells and then wish to isolate particular components from the cells. For example, a researcher may want to purify a protein that functions as a transcription factor. To do so, he or she would begin with a sample of cells that synthesize this protein and then break open the cells using one of the methods described in **Table A.1**. In eukaryotes, the breakage of cells releases the soluble proteins from the cell; it also dissociates the cell organelles that are bounded by membranes. This mixture of proteins and cell organelles can then be isolated and purified by centrifugation and chromatographic methods, which are described next.

Centrifugation

Centrifugation is a method commonly used to separate cell organelles and macromolecules. A **centrifuge** contains a motor that causes a rotor holding centrifuge tubes to spin very rapidly. As the rotor spins, particles move toward the bottom of the centrifuge tube; the rate at which they move depends on several factors, including their densities, sizes, and shapes, and the viscosity of the medium. The rate at which a macromolecule or cell

TABLE **A.1**

Common Methods of Cell Disruption

Method	Description
Sonication	The exposure of cells to intense sound waves, which breaks the cell membranes.
French press	The passage of cells through a small aperture under high pressure, which breaks the cell membranes and cell wall.
Homogenization	Cells are placed in a tube that contains a pestle. When the pestle is spun, the cells are squeezed through the small space between the pestle and the glass wall of the tube, thereby breaking them.
Osmotic shock	The transfer of cells into a hypo-osmotic medium. The cells take up water and eventually burst.

organelle sediments to the bottom of a centrifuge tube is called its **sedimentation coefficient,** which is normally expressed in Svedberg units (S). A sedimentation coefficient has the units of seconds: $1\ S = 1 \times 10^{-13}$ seconds.

When a sample contains a mixture of macromolecules or cell organelles, it is likely that different components will sediment at different rates. This phenomenon, known as **differential centrifugation,** is shown in **Figure A.3**. As seen here, particles with large sedimentation coefficients reach the bottom of the tube more quickly than those with smaller coefficients. Researchers can follow two different strategies that use differential centrifugation as a separation technique. One way is to separate the **supernatant** from the **pellet** following centrifugation. The pellet is a collection of particles found at the bottom of the tube, and the supernatant is the liquid found above the pellet. In Figure A.3, when the experimenter had subjected the sample to a low-speed spin, most of the particles with large sedimentation coefficients are found in the pellet, whereas most of the particles with small and intermediate coefficients are found in the supernatant. A high-speed spin of the supernatant then separates the small and intermediate particles. Therefore, differential centrifugation provides a way of segregating these three types of particles.

A second way to separate particles using centrifugation is to collect fractions. A **fraction** is a portion of the liquid contained within a centrifuge tube. The collection of fractions is done when the solution within the centrifuge tube contains a gradient. For example, as shown in **Figure A.4**, the solution at the top of the tube has a lower concentration of cesium chloride (CsCl) than that at the bottom. In this experiment, a sample is layered on the top of the gradient and then centrifuged. In this example, the DNA and RNA separate from each other, because they have different sedimentation coefficients. The experimenter then punctures the bottom of the tube and collects fractions. The DNA fragments, which are heavier, come out of the tube in the earlier fractions; the RNA molecules are collected in later fractions.

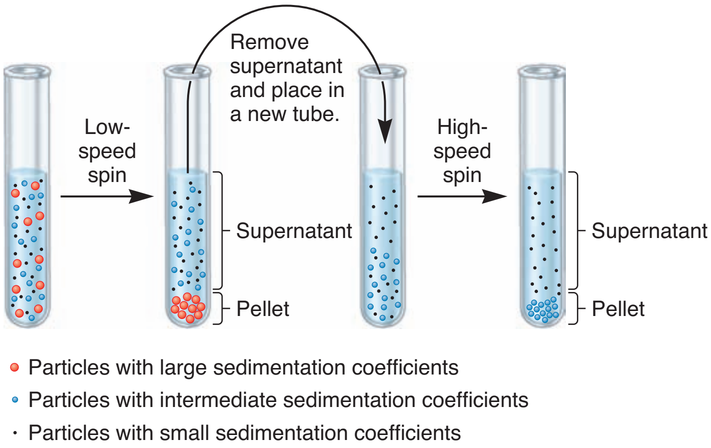

Low-
speed
spin
Remove
supernatant
and place in
a new tube.
High-
speed
spin
Supernatant
Pellet
Supernatant
Pellet
● Particles with large sedimentation coefficients
● Particles with intermediate sedimentation coefficients
· Particles with small sedimentation coefficients

FIGURE A.3 The method of differential centrifugation. A sample containing a mixture of particles with different sedimentation coefficients is placed in a centrifuge tube. The tube is subjected to a low-speed spin that pellets the particles with large sedimentation coefficients. After a high-speed spin of the supernatant, the particles with an intermediate sedimentation coefficient are found in the pellet, and those with a small sedimentation coefficient are in the liquid supernatant.

A type of gradient centrifugation that may also be used to separate macromolecules and organelles is **equilibrium density centrifugation.** In this method, the particles sediment through the gradient, reaching a position where the density of the particle matches the density of the solution. At this point, the particle is at equilibrium and does not move any farther toward the bottom of the tube.

Chromatography and Gel Electrophoresis

Chromatography is a method of separating different macromolecules and small molecules based on their chemical and physical properties. In this method, a sample is dissolved in a liquid

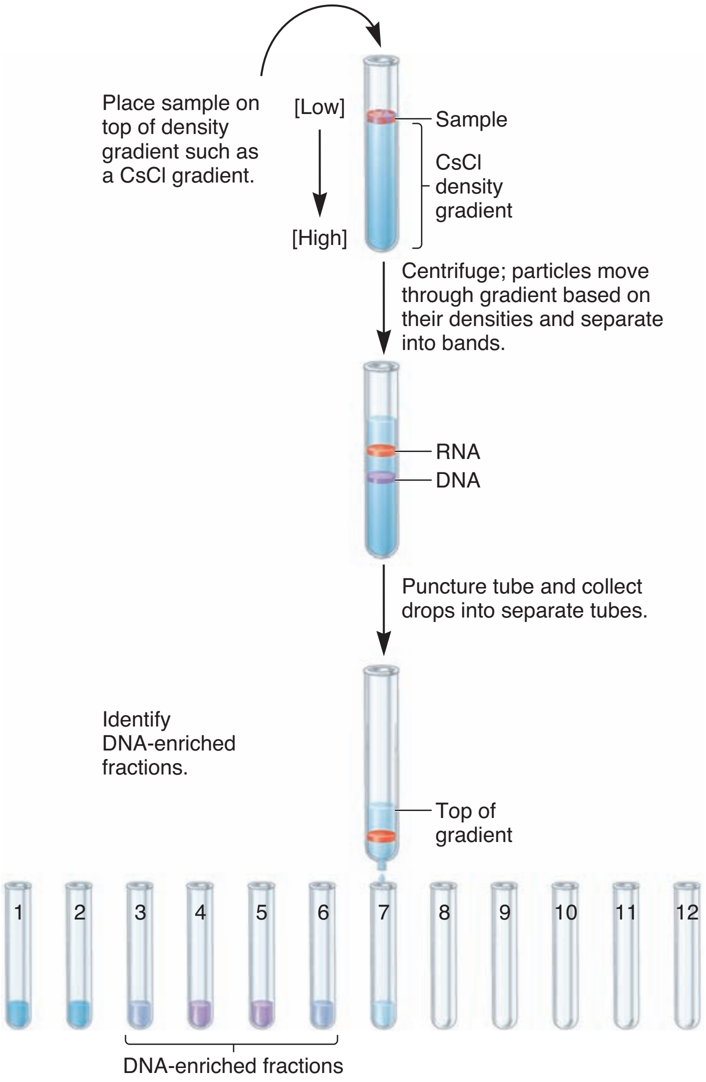

Place sample on
top of density
gradient such as
a CsCl gradient.
[Low]
[High]
Sample
CsCl
density
gradient
Centrifuge; particles move
through gradient based on
their densities and separate
into bands.
RNA
DNA
Puncture tube and collect
drops into separate tubes.
Identify
DNA-enriched
fractions.
Top of
gradient
1
2
3
4
5
6
7
8
9
10
11
12
DNA-enriched fractions

FIGURE A.4 Gradient centrifugation and the collection of fractions.

solvent and exposed to some type of matrix, such as a column containing beads or a thin strip of paper. The degree to which the molecules interact with the matrix depends on their chemical and physical characteristics. For example, a positively charged molecule binds tightly to a negatively charged matrix, but a neutral molecule does not.

Figure A.5 illustrates how column chromatography can be used to separate molecules that differ with respect to charge. Prior to this experiment, a column is packed with beads that are positively charged. There is plenty of space between the beads for molecules to flow from the top of the column to the bottom. However, if the molecules are negatively charged, they will spend some of their time binding to the positive charges on the surface of the beads. In the example shown in Figure A.5, the red proteins are positively charged and, therefore, flow rapidly from the top of the column to the bottom. They emerge in the fractions that are collected early in this experiment. The blue proteins, however, are negatively charged and tend to bind to the beads. The binding of the blue proteins to the beads can be disrupted by changing the ionic strength or pH of the solution that is added to the column. Eventually, the blue proteins will be eluted (i.e., leave the column) in later fractions.

Researchers use many variations of chromatography to separate molecules and macromolecules. The type shown in Figure A.5 is called ion-exchange chromatography, because its basis for separation depends on the charge of the molecules. In another type of column chromatography, known as gel filtration chromatography, the beads are porous. Small molecules are temporarily trapped within the beads, whereas large molecules flow between the beads. In this way, gel filtration separates molecules on the basis of size. To separate different types of macromolecules, such as proteins, researchers may use another type of bead; this bead has a preattached molecule that binds specifically to the protein they want to purify. For example, if a transcription factor binds a particular DNA sequence as part of its function, the beads within a column may have this DNA sequence preattached to them. Therefore, the transcription factor binds tightly to the DNA attached to these beads, whereas all other proteins are eluted rapidly from the column. This form of chromatography is called affinity chromatography, because the beads have a special affinity for the macromolecule of interest.

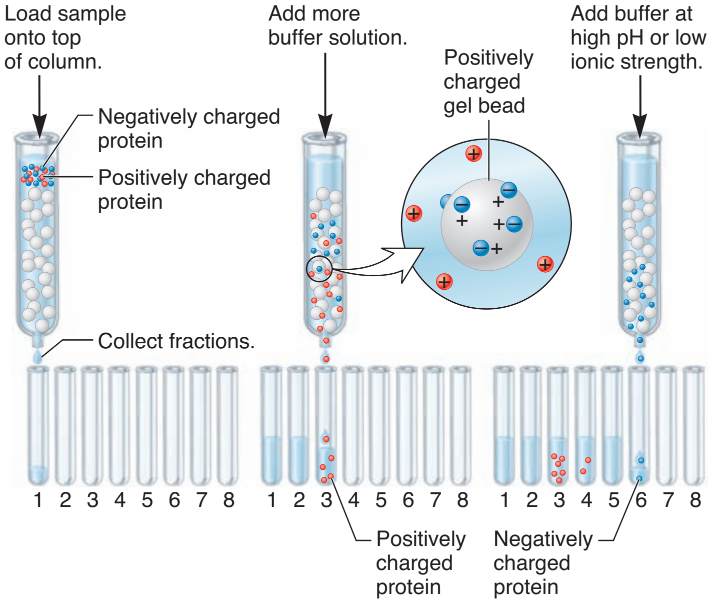

Load sample onto top of column.
Add more buffer solution.
Positively charged gel bead
Add buffer at high pH or low ionic strength.
Negatively charged protein
Positively charged protein
Collect fractions.
+
−
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
Positively charged protein
Negatively charged protein

FIGURE A.5 **Ion-exchange chromatography.**

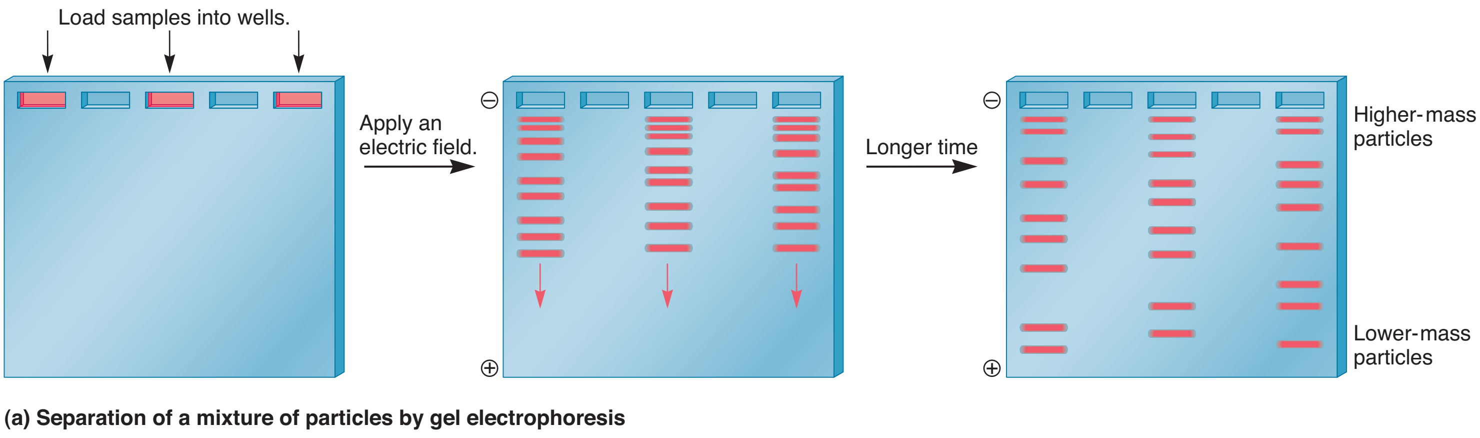
Besides column chromatography, in which beads are packed into a column, a matrix can be made in other ways. In paper chromatography, molecules pass through a matrix composed of paper. The rate of movement of molecules through the paper depends on their degree of interaction with the solvent and paper. In thin-layer chromatography, a matrix is spread out as a very thin layer on a rigid support such as a glass plate. In general, paper and thin-layer chromatography are effective at separating small molecules, whereas column chromatography is used to separate macromolecules such as DNA fragments or proteins.

Gel electrophoresis combines chromatography and electrophoresis to separate molecules and macromolecules. As its name suggests, the matrix used in gel electrophoresis is composed of a gel. As shown in **Figure A.6**, samples are loaded into wells at one end of the gel, and an electric field is applied across the gel. This electric field causes charged molecules to migrate from one side of the gel to the other. The migration of molecules in response to an electric field is called **electrophoresis.** In the examples of gel electrophoresis found in this textbook, the macromolecules within the sample migrate toward the positive end of the gel. In most forms of gel electrophoresis, a mixture of macromolecules is separated according to their molecular masses. Small proteins or DNA fragments move to the bottom of the gel more quickly than larger ones. Because the samples are loaded in rectangular wells at the top of the gel, the molecules within the sample are separated into bands within the gel. These bands of separated macromolecules can be visualized with stains. For example, ethidium bromide is a stain that binds to DNA and RNA and can be seen under ultraviolet light.

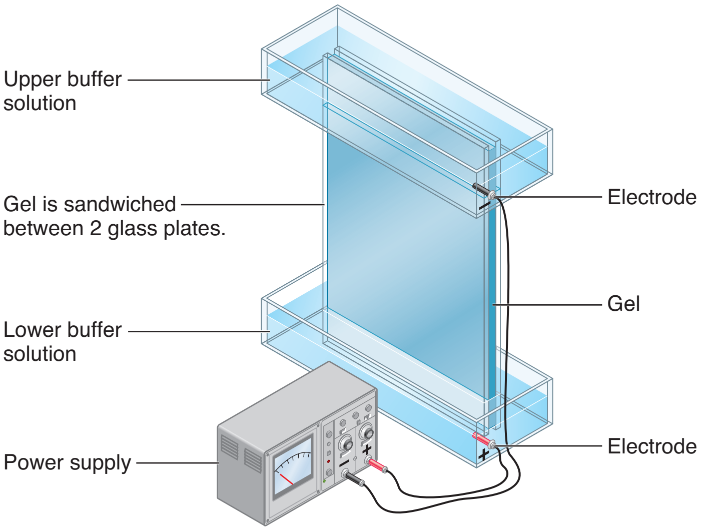
The two most commonly used gels are polymers made from acrylamide or agarose. Proteins typically are separated on polyacrylamide gels, whereas DNA fragments are separated on agarose gels. Occasionally, researchers use polyacrylamide gels to separate DNA fragments that are relatively small (namely, less than 1000 bp in length).

A.4 METHODS FOR MEASURING CONCENTRATIONS OF MOLECULES AND DETECTING RADIOISOTOPES AND ANTIGENS

To understand the structure and function of cells, researchers often need to detect the presence of molecules and macromolecules and to measure their concentrations. In this section, we consider a variety of methods for detecting and measuring the concentrations of biological molecules and macromolecules.


Load samples into wells.
Apply an electric field.
Longer time
Higher-mass particles
Lower-mass particles
⊖
⊕

(a) Separation of a mixture of particles by gel electrophoresis


Upper buffer solution
Gel is sandwiched between 2 glass plates.
Lower buffer solution
Power supply
Electrode
Gel
Electrode

(b) Apparatus used in gel electrophoresis

FIGURE A.6 **Acrylamide gel electrophoresis of DNA fragments.**

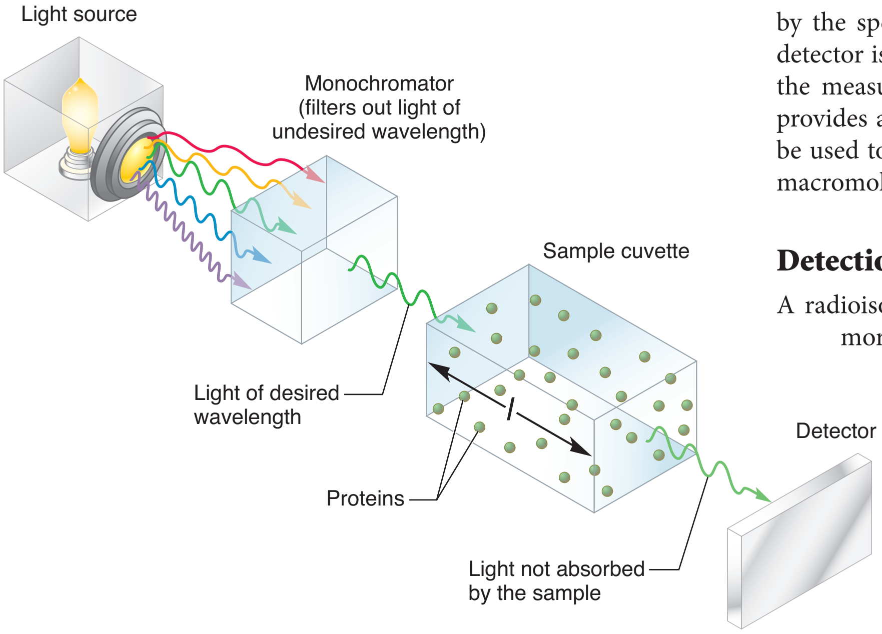

Light source
Monochromator (filters out light of undesired wavelength)
Sample cuvette
Light of desired wavelength
Proteins
l
Detector
Light not absorbed by the sample

FIGURE A.7 **Design of a spectrophotometer.**

Spectroscopy

Macromolecules found in living cells, such as proteins, DNA, and RNA, are fairly complex molecules that can absorb radiation (e.g., light). Likewise, small molecules such as amino acids and nucleotides can also absorb light. A device known as a **spectrophotometer** is used by researchers to determine how much radiation at various wavelengths a sample can absorb. The amount of absorption can be used to determine the concentration of particular molecules within a sample, because each type of molecule or macromolecule has its own characteristic wavelength(s) of absorption, called its absorption spectrum.

A spectrophotometer typically has two light sources, which can emit ultraviolet or visible light. As shown in Figure A.7, the light source is passed through a monochromator, which emits the light at a desired wavelength. This incident light then strikes a sample contained within a cuvette. Some of the incident light is absorbed, and some is not. The amount and wavelengths of light that are absorbed depend on the concentration and structures of the molecules and macromolecules in the cuvette. The unabsorbed light passes through the sample and is detected by the spectrophotometer. The amount of light that strikes the detector is subtracted from the amount of incident light, yielding the measure of absorption. In this way, the spectrophotometer provides an absorption reading for the sample. This reading can be used to calculate the concentration of particular molecules or macromolecules in a sample.

Detection of Radioisotopes

A radioisotope is an unstable form of an atom that decays to a more stable form by emitting α-, β-, or γ-rays, which are types of ionizing radiation. In research, radioisotopes that are β and/or γ emitters are commonly used. A β-ray is an emitted electron, and a γ-ray is an emitted photon. Some radioisotopes commonly used in biological experiments are shown in Table A.2.

Experimentally, radioisotopes are used because they are easy to detect. Therefore, if a particular compound is radiolabeled, its presence can be detected specifically throughout the course of the

TABLE **A.2**

Some Useful Isotopes in Genetics

Isotope	Stable or Radioactive	Emission	Half-life
^{2}H	Stable		
^{3}H	Radioactive	β	12.3 years
^{13}C	Stable		
^{14}C	Radioactive	β	5730 years
^{15}N	Stable		
^{18}O	Stable		
^{24}Na	Radioactive	β (and γ)	15 hours
^{32}P	Radioactive	β	14.3 days
^{35}S	Radioactive	β	87.4 days
^{45}Ca	Radioactive	β	164 days
^{59}Fe	Radioactive	β (and γ)	45 days
^{131}I	Radioactive	β (and γ)	8.1 days

experiment. For example, if a nucleotide is radiolabeled with ^{32}P, a researcher can determine whether the isotope becomes incorporated into newly made DNA or whether it remains as the free nucleotide. Researchers commonly use two different methods of detecting radioisotopes: scintillation counting and autoradiography.

The technique of **scintillation counting** permits a researcher to count the number of radioactive emissions from a sample containing a population of radioisotopes. In this approach, the sample is dissolved in a solution (called the scintillant) that contains organic solvents and one or more compounds known as fluors. When radioisotopes emit ionizing radiation, the energy is absorbed by the fluors in the solvent. This excites the fluor molecules, causing their electrons to be boosted to higher energy levels. The excited electrons return to lower, more stable energy levels by releasing photons of light. When a fluor is struck by ionizing radiation, it also absorbs the energy and then releases a photon of light within a particular wavelength range. The role of a device known as a scintillation counter is to count the photons of light emitted by the fluor. **Figure A.8** shows a scintillation counter. To use this device, a researcher dissolves her or his sample in a scintillant and then places the sample in a scintillation vial. The vial is then placed in the scintillation counter, which detects the amount of radioactivity. The scintillation counter has a digital meter that displays the amount of radioactivity in the sample and provides a printout of the amount of radioactivity in counts per minute. A scintillation counter contains several rows for the loading and analysis of many scintillation vials. After they have been loaded, the scintillation counter counts the amount of radioactivity in each vial and provides the researcher with a printout of the amount of radioactivity in each vial.

A second way of detecting radioisotopes is via **autoradiography.** This technique is not as quantitative as scintillation counting, because it does not provide the experimenter with a precise measure of the amount of radioactivity in counts per minute. However, autoradiography has the great advantage that it can detect the location of radioisotopes as they are found in macromolecules or cells. For example, autoradiography is used to detect a particular band on a gel or to map the location of a gene within an intact chromosome.

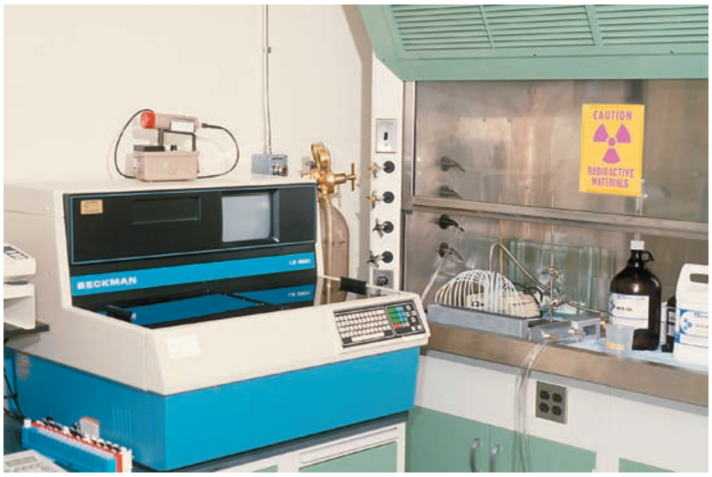


FIGURE A.8 **A scintillation counter.**

To conduct autoradiography, a sample containing a radioisotope is fixed and usually dried. If it is a cellular sample, it also may be thin-sectioned. The sample is then pressed next to X-ray film (in the dark) and placed in a lightproof cassette. When a radioisotope decays, it emits a β- or γ-ray, which may strike a thin layer of photoemulsion next to the film. The photoemulsion contains silver salts such as AgBr. When a radioactive particle is emitted and strikes the photoemulsion, a silver grain is deposited on the film. This produces a dark spot on the film, which correlates with the original location of the radioisotope in the sample. In this way, the dark image on the film reveals the location(s) of the radioisotopes in the sample. Figure 11.4b shows how autoradiography can be used to visualize the process of bacterial chromosome replication. In this case, radiolabeled nucleotides were incorporated into the DNA, making it possible to picture the topology of the chromosome as two replication forks pass around the circular chromosome.

ANSWERS TO CONCEPT CHECKS AND EVEN-NUMBERED PROBLEMS

B

CHAPTER 1

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 1.1 Understanding our genes may help in the diagnosis of inherited diseases. It may also lead to the development of drugs to combat diseases. Other answers are possible.

FIGURE 1.2 Many ethical issues are associated with human cloning. Is it the wrong thing to do? Does it conflict with an individual's religious views? And so on.

FIGURE 1.3 By sorting the mosquitoes, sterile males can be released into the environment to limit mosquito reproduction, because females mate only once.

FIGURE 1.4 DNA is a macromolecule.

FIGURE 1.5 DNA and proteins are found in chromosomes. A small amount of RNA may also be associated with chromosomes when transcription is occurring.

FIGURE 1.6 The information to make a polypeptide is stored in DNA.

FIGURE 1.7 The dark-colored butterfly has a more active pigment-producing enzyme.

FIGURE 1.8 Genetic variation is the reason these frogs look different.

FIGURE 1.9 These are examples of variation in chromosome number.

FIGURE 1.10 The girl in the center was given a phenylalanine-free diet, which prevented the harmful symptoms of PKU.

FIGURE 1.11 A corn gamete would contain 10 chromosomes. (The leaf cells are diploid.)

FIGURE 1.12 The horse populations have become adapted to their environment, which has changed over the course of many years.

FIGURE 1.13 There are several possible examples of other model organisms, including rats and frogs.

End-of-Chapter Questions:

Conceptual Questions

C2. The structure and function of proteins govern the structure and function of living cells. The cells of the body determine an organism's traits.

C4. A. Molecular level. This is a description of a how an allele affects protein function.

B. Cellular level. This is a description of how protein function affects cell structure.

C. Population level. This is a description of how the two alleles affect members of a population.

D. Organism level. This is a description of how the alleles affect the traits of an individual.

C6. An extra chromosome (specifically an extra copy of chromosome 21) causes Down syndrome.

C8. The term *diploid* means that a cell has two copies of each type of chromosome. In humans, nearly all of the cells are diploid except for gametes (i.e., sperm and egg cells). Gametes usually have only one set of chromosomes.

C10. The genetic code is the way in which the sequence of bases in RNA is read to produce a sequence of amino acids within a protein.

C12. The statement in part A is not correct. Individuals do not evolve. Populations evolve because certain individuals are more likely to survive and reproduce and pass their genes to succeeding generations.

Application and Experimental Questions

E2. A genetic cross involves breeding two different individuals.

E4. You would see 47 chromosomes instead of 46. There would be three copies of chromosome 21 instead of two copies.

E6. You need to follow the scientific method. You can take a look at an experiment in another chapter to see how the scientific method is followed.

CHAPTER 2

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 2.1 Compartmentalization means that cells have membrane-bound compartments.

FIGURE 2.2 The chromosomes would not be spread out very well, and would probably be overlapping. It would be difficult to see individual chromosomes.

FIGURE 2.3 Homologs are similar in size and banding pattern, and they carry the same types of genes. However, the alleles of a given gene may be different.

FIGURE 2.4 FtsZ assembles into a ring at the future site of the septum and recruits other proteins to this site that produce a cell wall between the two daughter cells.

FIGURE 2.5 The G_1 phase is a phase of the cell cycle during which a cell may enter into the cell division process. By comparison, the G_0 phase is a phase in which a cell is either not progressing through the cell cycle or will never divide again.

FIGURE 2.6 Homologs are genetically similar; one is inherited from the mother and the other from the father. By comparison, chromatids are the product of DNA replication. The chromatids within a pair of sister chromatids are genetically identical.

FIGURE 2.7 One end of a kinetochore microtubule is attached to a kinetochore on a chromosome. The other end is within the centrosome.

FIGURE 2.8 Anaphase.

FIGURE 2.9 Ingression occurs because myosin motor proteins shorten the contractile ring, which is formed from actin proteins.

FIGURE 2.10 The end result of crossing over is that homologous chromosomes have exchanged pieces.

FIGURE 2.11 The synaptonemal complex is between homologs.

FIGURE 2.12 The cells at the end of meiosis are haploid, whereas the mother cell is diploid.

FIGURE 2.13 In mitosis, each pair of sister chromatids is attached to both poles, whereas in metaphase of meiosis I, each pair of sister chromatids is attached to just one pole.

FIGURE 2.14 Polar bodies are small cells produced during oogenesis that degenerate.

FIGURE 2.15 All of the nuclei in the embryo sac are haploid. The central cell has two haploid nuclei and all of the other cells, including the egg, have just one.

End-of-Chapter Questions:

Conceptual Questions

C2. The term *homolog* refers to the members of a chromosome pair. Homologs are usually the same size and carry the same types and order of genes. They may differ in that the genes they carry may be different alleles.

C4. Metaphase is the organization phase and anaphase is the separation phase.

C6. In metaphase I of meiosis, each pair of chromatids is attached to only one pole via the kinetochore microtubules. In metaphase of mitosis, there are two attachments (i.e., to both poles). If the attachment was lost, a chromosome would not migrate to a pole and may not become enclosed in a nuclear membrane after telophase. If left out in the cytoplasm, it would eventually be degraded.

C8. The reduction occurs because there is a single DNA replication event but two cell divisions. Because of the nature of separation during anaphase I, each cell receives one copy of each type of chromosome.

C10. It means that the maternally derived and paternally derived chromosomes are randomly aligned along the metaphase plate during metaphase I.

C12. There are three pairs of chromosomes. The number of different, random alignments equals 2^n, where n equals the number of chromosomes per set. So the possible number of arrangements equals 2^3, which is 8.

C14. The probability would be much lower because pieces of maternal chromosomes would be mixed with the paternal chromosomes. Therefore, it is unlikely to inherit a chromosome that was completely paternally derived.

C16. During interphase, the chromosomes are greatly extended. In this conformation, they might get tangled up with each other and not sort properly during meiosis and mitosis. The condensation process probably occurs so that the chromosomes easily align along the equatorial plate during metaphase without getting tangled up.

C18. During prophase II, your drawing should show four replicated chromosomes (i.e., four structures that look like Xs). Each chromosome is one homolog. During prophase of mitosis, there should be eight replicated chromosomes (i.e., eight Xs). During prophase of mitosis, there are pairs of homologs. The main difference is that prophase II has a single copy of each of the four chromosomes, whereas prophase of mitosis has four pairs of homologs. At the end of meiosis I, each daughter cell has received only one copy of a homologous pair, not both. This is due to the alignment of homologs during metaphase I and their separation during anaphase I.

C20. DNA replication does not take place during interphase II. The chromosomes at the end of telophase I have already replicated (i.e., they are found in pairs of sister chromatids). During meiosis II, the sister chromatids separate from each other, yielding individual chromosomes.

C22. A. 20 B. 10 C. 30 D. 20

C24. Male gametes are usually small and mobile. Animal and some plant male gametes often contain flagella, which make them motile. The mobility of the male gamete enables it to come in contact with the female gamete. Female gametes are usually much larger and nonmotile and contain nutrients to help the growth of the embryo after fertilization occurs.

C26. During oogenesis in humans, the cells are arrested in prophase I of meiosis for many years until selected primary oocytes progress through the rest of meiosis I and begin meiosis II. If fertilization occurs, meiosis II is completed.

Application and Experimental Questions

E2. During interphase, the chromosomes are longer, thinner, and much harder to see. In metaphase, they are highly condensed, which makes them thicker and shorter.

Questions for Student Discussion/Collaboration

2. A major advantage of sexual reproduction is that it fosters genetic diversity within future populations. A major disadvantage is that individuals of opposite sex must find a mate.

CHAPTER 3

Note: the answers to Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 3.2 The male gamete is found within pollen grains.

FIGURE 3.3 The white flower is providing the sperm, and the purple flower is providing the eggs.

FIGURE 3.4 A true-breeding strain maintains the same trait over the course of many generations.

FIGURE 3.6 Segregation means that the *T* and *t* alleles separate from each other so that a gamete receives one of them, but not both.

FIGURE 3.7 In this hypothesis, two different genes are linked. The alleles of the same gene are not linked.

FIGURE 3.9 Independent assortment allows for new combinations of alleles among different genes to be found in future generations of offspring.

FIGURE 3.10 Such a parent could make two types of gametes, *Ty* and *ty*, in equal proportions.

FIGURE 3.11 Homologous chromosomes separate at anaphase of meiosis I.

FIGURE 3.12 These chromosomes could line up in four different ways.

FIGURE 3.13 Horizontal lines connect two individuals that have offspring together, and they connect all of the offspring that are produced by the same two parents.

End-of-Chapter Questions:

Conceptual Questions

C2. In the case of plants, cross-fertilization occurs when the pollen and eggs come from different plants while in self-fertilization they come from the same plant.

C4. A homozygote that has two copies of the same allele.

C6. Diploid organisms contain two copies of each type of gene. When they make gametes, only one copy of each gene is found in a gamete. Two alleles cannot stay together within the same gamete.

C8. Genotypes: 1:1 *Tt* and *tt*

Phenotypes: 1:1 Tall and dwarf

C10. *c* is the recessive allele for constricted pods; *Y* is the dominant allele for yellow color. The cross is *ccYy* × *CcYy*. Follow the directions for setting up a Punnett square, as described in Chapter 3. The genotypic ratio is

$$2\ CcYY : 4\ CcYy : 2\ Ccyy : 2\ ccYY : 4\ ccYy : 2\ ccyy$$

This 2:4:2:2:4:2 ratio could be reduced to a 1:2:1:1:2:1 ratio.

The phenotypic ratio is 6 smooth pods, yellow seeds : 2 smooth pods, green seeds: 6 constricted pods, yellow seeds : 2 constricted pods, green seeds. This 6:2:6:2 ratio could be reduced to a 3:1:3:1 ratio.

C12. Offspring with a nonparental phenotype are consistent with the idea of independent assortment. If two different traits were always transmitted together as unit, it would be impossible to get nonparental phenotypic combinations. For example, if a true-breeding parent had two dominant traits and was crossed to a true-breeding parent having the two recessive traits, the F_2 generation could not have offspring with one recessive and one dominant phenotype. However, because independent assortment can occur, it is possible for F_2 offspring to have one dominant and one recessive trait.

C14. A. Barring a new mutation during gamete formation, the chance is 100% because they must be heterozygotes in order to produce a child with a recessive disorder.

B. Construct a Punnett square. There is a 50% chance of heterozygous children.

C. Use the product rule. The chance of being phenotypically normal is 0.75 (i.e., 75%), so the answer is $0.75 \times 0.75 \times 0.75 = 0.422$, which is 42.2%.

D. Use the binomial expansion equation where $n = 3$, $x = 2$, $p = 0.75$, $q = 0.25$. The answer is 0.422, or 42.2%.

C16. First construct a Punnett square. The chances are 75% of producing a solid-colored pup and 25% of producing a spotted pup.

A. Use the binomial expansion equation where $n = 5$, $x = 4$, $p = 0.75$, $q = 0.25$. The answer is 0.396 = 39.6% of the time.

B. You can use the binomial expansion equation for each litter. For the first litter, $n = 6$, $x = 4$, $p = 0.75$, $q = 0.25$; for the second litter, $n = 5$, $x = 5$, $p = 0.75$, $q = 0.25$. Because the litters are in a specified order, we use the product rule and multiply the probability of the first litter times the probability of the second litter. The answer is 0.070, or 7.0%.

C. To calculate the probability of the first litter, we use the product rule and multiply the probability of the first pup (0.75) times the probability of the remaining four. We use the binomial expansion equation to calculate the probability of the remaining four, where $n = 4$, $x = 3$, $p = 0.75$, $q = 0.25$. The probability of the first litter is 0.316. To calculate the probability of the second litter, we use the product rule and multiply the probability of the first pup (0.25) times the probability of the second pup (0.25) times the probability of the remaining five. To calculate the probability of the remaining five, we use the binomial expansion equation, where $n = 5$, $x = 4$, $p = 0.75$, $q = 0.25$. The probability of the second litter is 0.025. To get the probability of these two litters occurring in this order, we use the product rule and multiply the probability of the first litter (0.316) times the probability of the second litter (0.025). The answer is 0.008, or 0.8%.

D. Because this is a specified order, we use the product rule and multiply the probability of the firstborn (0.75) times the probability of the second born (0.25) times the probability of the remaining four. We use the binomial expansion equation to calculate the probability of the remaining four pups, where $n = 4$, $x = 2$, $p = 0.75$, $q = 0.25$. The answer is 0.040, or 4.0%.

C18. A. Use the product rule: $(1/4)(1/4)=1/16$

B. Use the binomial expansion equation:

$n = 4, p = {}^{1}/_{4}, q = {}^{3}/_{4}, x = 2$

$P = 0.21 = 21\%$

C. Use the product rule:

$$(1/4)(3/4)(3/4) = 0.14, \text{ or } 14\%$$

C20. A. 1/4 B. 1, or 100%

C. $(3/4)(3/4)(3/4) = 27/64 = 0.42$, or 42%

D. Use the binomial expansion equation where

$$n = 7, p = 3/4, q = 1/4, x = 3$$

$$P = 0.058, \text{ or } 5.8\%$$

E. The probability that the first plant is tall is 3/4. To calculate the probability that among the next four, any two will be tall, we use the binomial expansion equation, where $n = 4$, $p = 3/4$, $q = 1/4$, and $x = 2$.

The probability P equals 0.21.

To calculate the overall probability of these two events:

$$(3/4)(0.21) = 0.16, \text{ or } 16\%$$

C22. It violates the law of segregation because two copies of one gene are in the gamete. The two alleles for the *A* gene did not segregate from each other.

C24. Based on this pedigree, it is likely to be dominant inheritance because an affected child always has an affected parent. In fact, it is a dominant disorder.

C26. It is impossible for the F_1 individuals to be true-breeding because they are all heterozygotes.

C28. 2 *TY*, *tY*, 2 *Ty*, *ty*, *TTY*, *TTy*, 2 *TtY*, 2 *Tty*

It may be tricky to think about, but you get 2 *TY* and 2 *Ty* because either of the two *T* alleles could combine with *Y* or *y*. Also, you get 2 *TtY* and 2 *Tty* because either of the two *T* alleles could combine with *t* and then combine with *Y* or *y*.

C30. The genotype of the F_1 plants is *Tt Yy Rr*. According to the laws of segregation and independent assortment, the alleles of each gene will segregate from each other, and the alleles of different genes will randomly assort into gametes. A *Tt Yy Rr* individual could make eight types of gametes: *TYR*, *TyR*, *Tyr*, *TYr*, *tYR*, *tyR*, *tYr*, and *tyr*, in equal proportions (i.e., 1/8 of each type of gamete). To determine genotypes and phenotypes, you could make a large Punnett square that would contain 64 boxes. You would need to line up the eight possible gametes across the top and along the side, and then fill in the 64 boxes. Alternatively, you could use one of the two approaches described in Solved Problem S3. The genotypes and phenotypes would be:

1 *TT YY RR*
2 *TT Yy RR*
2 *TT YY Rr*
2 *Tt YY RR*
4 *TT Yy Rr*
4 *Tt Yy RR*
4 *Tt YY Rr*
8 *Tt Yy Rr* = 27 tall, yellow, round
1 *TT yy RR*
2 *Tt yy RR*
2 *TT yy Rr*
4 *Tt yy Rr* = 9 tall, green, round
1 *TT YY rr*
2 *TT Yy rr*
2 *Tt YY rr*
4 *Tt Yy rr* = 9 tall, yellow, wrinkled
1 *tt YY RR*
2 *tt Yy RR*
2 *tt YY Rr*

4 *tt Yy Rr* = 9 dwarf, yellow, round
1 *TT yy rr*
2 *Tt yy rr* = 3 tall, green, wrinkled
1 *tt yy RR*
2 *tt yy Rr* = 3 dwarf, green, round
1 *tt YY rr*
2 *tt Yy rr* = 3 dwarf, yellow, wrinkled
1 *tt yy rr* = 1 dwarf, green, wrinkled

C32. The wooly-haired male is a heterozygote, because he has the trait and his mother did not. (He must have inherited the normal allele from his mother.) Therefore, he has a 50% chance of passing the wooly allele to his offspring; his offspring have a 50% of passing the allele to their offspring; and these grandchildren have a 50% chance of passing the allele to their offspring (the wooly haired man's great-grandchildren). Because this is an ordered sequence of independent events, we use the product rule: 0.5 × 0.5 × 0.5 = 0.125, or 12.5%. Because no other Scandinavians are on the island, the chance is 87.5% for the offspring being normal (because they could not inherit the wooly hair allele from anyone else). We use the binomial expansion equation to determine the likelihood that one out of eight great-grandchildren will have wooly hair, where $n = 8$, $x = 1$, $p = 0.125$, $q = 0.875$. The answer is 0.393, or 39.3%, of the time.

C34. Use the product rule. If the woman is heterozygous, there is a 50% chance of having an affected offspring: $(0.5)^7 = 0.0078$, or 0.78%, of the time. This is a pretty small probability. If the woman has an eighth child who is unaffected, however, she has to be a heterozygote, because it is a dominant trait. She would have to pass a normal allele to an unaffected offspring. The answer is 100%.

Application and Experimental Questions

E2. The experimental difference depends on where the pollen comes from. In self-fertilization, the pollen and eggs come from the same plant. In cross-fertilization, they come from different plants.

E4. According to Mendel's law of segregation, the genotypic ratio should be 1 homozygote dominant : 2 heterozygotes : 1 homozygote recessive. This data table considers only the plants with a dominant phenotype. The genotypic ratio should be 1 homozygote dominant : 2 heterozygotes. The homozygote dominants would be true-breeding, but the heterozygotes would not be true-breeding. This 1:2 ratio is very close to what Mendel observed.

E6. All three offspring had black fur. The ovaries from the albino female could only produce eggs with the dominant black allele (because they were obtained from a true-breeding black female). The actual phenotype of the albino mother does not matter. Therefore, all offspring would be heterozygotes (*Bb*) and have black fur.

E8. If we construct a Punnett square according to Mendel's laws, we expect a 9:3:3:1 ratio. Because a total of 556 offspring were observed, the expected number of offspring are

556 × 9/16 = 313 round, yellow

556 × 3/16 = 104 wrinkled, yellow

556 × 3/16 = 104 round, green

556 × 1/16 = 35 wrinkled, green

If we plug the observed and expected values into the chi square equation, we get a value of 0.51. With four categories, our degrees of freedom equal $n - 1$, or 3. If we look up the value of 0.51 in the chi square table (see Table 3.2), we see that it falls between the *P* values of 0.80 and 0.95. This means that the probability is 80% to 95% that any deviation between observed results and expected results was caused by random sampling error. Therefore, we accept the hypothesis. In other words, the results are consistent with the law of independent assortment.

E10. A. If we let c^+ represent normal wings and c represent curved wings, and e^+ represents gray body and e represents ebony body:

Parental cross: $cce^+e^+ \times c^+c^+ee$.

F_1 generation is heterozygous c^+ce^+e

An F_1 offspring crossed to flies with curved wings and ebony bodies is

$$c^+ce^+e \times ccee$$

The F_2 offspring would be a 1:1:1:1 ratio of flies:

$$c^+ce^+e : c^+cee : cce^+e : ccee$$

B. The phenotypic ratio of the F_2 flies would be a 1:1:1:1 ratio of flies:

normal wings, gray body : normal wings, ebony bodies : curved wings, gray bodies : curved wings, ebony bodies

C. From part B, we expect 1/4 of each category. There are a total of 444 offspring. The expected number of each category is 1/4 × 444, which equals 111.

$$\chi^2 = \frac{(114 - 111)^2}{111} + \frac{(105 - 111)^2}{111} + \frac{(111 - 111)^2}{111} + \frac{(114 - 111)^2}{111}$$

$$\chi^2 = 0.49$$

With 3 degrees of freedom, a value of 0.49 or greater is likely to occur between 80% and 95% of the time. Therefore, we accept our hypothesis.

E12. Follow through the same basic chi square strategy as before. We expect a 3:1 ratio, or 3/4 of the dominant phenotype and 1/4 of the recessive phenotype.

The observed and expected values are as follows (rounded to the nearest whole number):

Observed*	Expected	$\frac{(O - E)^2}{E}$
5474	5493	0.066
1850	1831	0.197
6022	6017	0.004
2001	2006	0.012
705	697	0.092
224	232	0.276
882	886	0.018
299	295	0.054
428	435	0.113
152	145	0.338
651	644	0.076
207	215	0.298
787	798	0.152
277	266	0.455
		$\chi^2 = 2.15$

*Due to rounding, the observed and expected values may not add up to precisely the same number.

Because $n = 14$, there are 13 degrees of freedom. If we look up this value in the chi square table, we have to look between 10 and 15 degrees of freedom. In either case, we would expect the value of 2.15 or greater to occur more than 99% of the time. Therefore, we accept the hypothesis.

E14. The dwarf parent with terminal flowers must be homozygous for both genes, because it is expressing these two recessive traits: *ttaa*, where *t* is the recessive dwarf allele, and *a* is the recessive allele for terminal flowers. The phenotype of the other parent is dominant for both traits. However, because this parent was able to produce dwarf offspring with axial flowers, it must have been heterozygous for both genes: *TtAa*.

E16. Our hypothesis is that blue flowers and purple seeds are dominant traits and they are governed by two genes that assort independently. According to this hypothesis, the F_2 generation should yield a ratio of 9 blue flowers, purple seeds : 3 blue flowers, green seeds : 3 white flowers, purple seeds : 1 white flower, green seeds. Because there are a total of 300 offspring produced, the expected numbers would be

9/16 × 300 = 169 blue flowers, purple seeds

3/16 × 300 = 56 blue flowers, green seeds

3/16 × 300 = 56 white flowers, purple seeds

1/16 × 300 = 19 white flowers, green seeds

If we look up this value in the chi square table under 3 degrees of freedom, the value is much higher than would be expected 1% of the time by chance alone. Therefore, we reject the hypothesis. The idea that the two genes are assorting independently seems to be incorrect. The F_1 generation supports the idea that blue flowers and purple seeds are dominant traits.

Questions for Student Discussion/Collaboration

2. If we construct a Punnett square, the following probabilities will be obtained:

tall with axial flowers 3/8

dwarf with terminal flowers 1/8

To calculate the probability of being tall with axial flowers or dwarf with terminal flowers, we use the sum rule:

$$3/8 + 1/8 = 4/8 = 1/2$$

We use the product rule to calculate the ordered events of the first three offspring being tall/axial or dwarf/terminal, and fourth offspring being tall axial:

$$(1/2)(1/2)(1/2)(3/8) = 3/64 = 0.047 = 4.7\%$$

CHAPTER 4

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 4.1 In the X-Y system, the presence of the Y chromosome causes maleness, whereas in the X-0 system, it is the ratio between the number of X chromosomes and number of sets of autosomes that determines sex. A ratio of 0.5 is male and 1.0 is female.

FIGURE 4.2 A male bee is not produced by sexual reproduction because it is produced from an unfertilized egg. Sexual reproduction involves the union of gametes.

FIGURE 4.3 A high temperature would favor a high percentage of female alligators. This might limit population size if there are insufficient numbers of males for reproduction. On the other, a high number of females might favor a larger population size if males are not limiting.

FIGURE 4.4 The sporophytes are the opposite sexes in dioecious plants.

FIGURE 4.5 The Barr body is more brightly staining because it is very compact.

FIGURE 4.6 X inactivation initially occurs during embryonic development.

FIGURE 4.7 Only the maintenance phase occurs in an adult female.

FIGURE 4.8 Homologous regions of the X and Y chromosome are important for chromosome synapsis (pairing) during meiosis.

FIGURE 4.9 In the F_2 generation, only the males had white eyes.

FIGURE 4.10 The key pedigree feature that points to X-linked inheritance is that only males are affected with the disorder. Also, carrier females often have affected brothers.

FIGURE 4.11 The reciprocal cross yields a different result because females carry two copies of an X-linked gene, but males have only one.

End-of-Chapter Questions:

Conceptual Questions

C2. A. Dark males and light females; reciprocal: all dark offspring

B. All dark offspring; reciprocal: dark females and light males

C. All dark offspring; reciprocal: dark females and light males

D. All dark offspring; reciprocal: dark females and light males

C4. A. Female; there are no Y chromosomes.

B. Female; there are no Y chromosomes.

C. Male; the Y chromosome determines maleness.

D. Male; the Y chromosome determines maleness.

C6. A Barr body is a mammalian X chromosome that is highly condensed. It is found in somatic cells with two or more X chromosomes. Most genes on the Barr body are inactive.

C8. X inactivation in heterozygous females produces a mosaic pattern of gene expression. During early embryonic development, some cells have the maternal X chromosome inactivated and other cells have the paternal X chromosome inactivated; these embryonic cells will divide and produce billions of cells. In the case of a female that is heterozygous for a gene that affects pigmentation of the fur, this produces a variegated pattern of coat color. Because it is a random process in any given animal, two female cats will vary in where the orange and black patches occur. A variegated coat pattern could not occur in female marsupials due to X inactivation because the paternal X chromosome is always inactivated in the somatic cells of females.

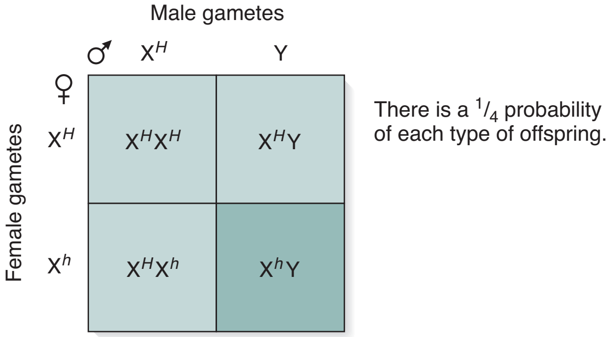
C10. The male is XXY. The person is male due to the presence of the Y chromosome. Because two X chromosomes are counted, one of the X chromosomes is inactivated to produce a Barr body.

C12. A. In females, one of the X chromosomes is inactivated. When the X chromosome that is inactivated carries the normal allele, only the defective color blindness allele will be expressed. Therefore, on average, about half of a female's eye cells are expected to express the normal allele. Depending on the relative amounts of cells expressing the normal versus the color-blind allele, the end result may be partial color blindness.

B. In this female, as a matter of chance, X inactivation occurred in the right eye to always, or nearly always, inactivate the X chromosome carrying the normal allele. The opposite occurred in the left eye. In the left eye, the chromosome carrying the color blindness allele was primarily inactivated.

C14. The spreading stage is when the X chromosome is inactivated (i.e., condensed) as a wave that spreads outward from the X-inactivation center (Xic). The condensation spreads from Xic to the rest of the X chromosome. The *Xist* gene is transcribed from the inactivated X chromosome. It encodes an RNA that coats the X chromosome, which subsequently attracts proteins that are responsible for the compaction.

C16. First set up the following Punnett square:



♂ / ♀	X^H	Y
X^H	X^HX^H	X^HY
X^h	X^HX^h	X^hY

There is a $^1/_4$ probability of each type of offspring.

A. 1/4 B. (3/4)(3/4)(3/4)(3/4) = 81/256 C. 3/4

D. The probability of an affected offspring is 1/4, and the probability of an unaffected offspring is 3/4. For this problem, you use the binomial expansion equation where $x = 2$, $n = 5$, $p = 1/4$, and $q = 3/4$. The answer is 0.26, or 26%, of the time.

Application and Experimental Questions

E2. In general, you cannot distinguish between autosomal and pseudoautosomal inheritance from a pedigree analysis. Mothers and fathers have an equal probability of passing the alleles to sons and daughters. However, if an offspring had a chromosomal abnormality, you might be able to tell. For example, in a family tree involving the *Mic2* allele, an offspring that was X0 would have less of the gene product and an offspring that was XXX or XYY or XXY would have extra amounts of the gene products. This may lead you to suspect that the gene is located on the sex chromosomes.

E4. Actually, his data are consistent with this hypothesis. To rule out a Y-linked allele, he could have crossed an F_1 female with a red-eyed male rather than an F_1 male. The same results would be obtained. Because the red-eyed male would not have a white allele, this would rule out Y-linkage.

E6. To be a white-eyed female, a fly must inherit two X chromosomes and both must carry the white-eye allele. This could occur only if both X chromosomes in the female stayed together and the male gamete contained a Y chromosome. The white-eyed females would be XXY. To produce a red-eyed male, a female gamete lacking any sex chromosomes could unite with a normal male gamete carrying the X^{w+}. This would produce an X0, red-eyed male.

E8. These rare female flies would be XXY. Both X chromosomes would carry the white allele and the miniature allele. These female flies would have miniature wings because they would have inherited both X chromosomes from their mother.

Questions for Student Discussion/Collaboration

2. One possibility is fertilization of an abnormal female gamete. The white-eyed male parent could make sperm carrying the X chromosome that could fertilize a female gamete without any sex chromosomes. This would produce an X0, white-eyed male. This is the most likely explanation.

   Another possibility is that Morgan's stocks were not completely true-breeding. The allele creating the white phenotype may have occurred several generations earlier in a female fly. Perhaps, among the many red-eyed sisters, one of them could have already been a heterozygote. The rare sister could produce white-eyed male offspring.

   New mutations are a third, but unlikely possibility, because we already know that the mutation rate is very low. It took him two years to get one white-eyed male.

CHAPTER 5

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 5.1 Both of these colors are considered wild type because both of them are prevalent in natural populations.

FIGURE 5.2 Yes. The *PP* homozygote probably makes twice of the protein that is needed for purple pigment formation.

FIGURE 5.3 Individual III-2 shows incomplete penetrance.

FIGURE 5.4 Genes and the environment determine an organism's traits.

FIGURE 5.5 Fifty percent of the functional enzyme is not enough to give a red color.

FIGURE 5.6 It is often easier to observe incomplete dominance at the molecular/cellular level.

FIGURE 5.7 In this case, the heterozygote is resistant to malaria.

FIGURE 5.8 The scenario shown in part (a) explains the effects of the sickle cell allele.

FIGURE 5.9 The *i* allele is a loss-of-function allele.

FIGURE 5.10 You can't really be sure if he inherited the baldness allele from his mother or father. If a bald father is a heterozygote, he could pass either the bald or nonbald allele to his son. The mother may or may not have passed the bald allele to her son.

FIGURE 5.11 A heterozygous woman would not become bald.

FIGURE 5.12 Certain traits are expressed only in males or females, possibly due to differences in the levels of sex hormones or other factors that differ between the sexes.

FIGURE 5.13 The heterozygote has one normal copy of the gene, which allows for development to proceed in a way that is not too far from normal. Having two mutant copies of the gene probably messes up development to a degree that is incompatible with survival.

FIGURE 5.14 Epistasis means that the alleles of one gene mask the phenotypic effects of the alleles of a different gene. Complementation occurs when two strains exhibiting the same recessive trait produce offspring that show the dominant (wild-type) trait. This usually means the alleles for the recessive trait are in two different genes.

FIGURE 5.15 In some cases, a single-gene knockout does not have an effect due to gene redundancy. Other explanations are also possible.

FIGURE 5.16 The two genes are redundant. Having one normal copy of either gene produces a triangular capsule. If both genes are inactive, an ovate capsule is produced.

End-of-Chapter Questions:

Conceptual Questions

C2. Sex-influenced traits are influenced by the sex of the individual even though the gene that governs the trait may be autosomally inherited. Pattern baldness in people is an example. Sex-limited traits are an extreme example of sex influence. The expression of a sex-limited trait is limited to one sex. For example, colorful plumage in certain species of birds is limited to the male sex. Sex-linked traits involve traits whose genes are found on the sex chromosomes. Examples in humans include hemophilia and color blindness.

C4. If the normal allele is dominant, it tells you that one copy of the gene produces a sufficient amount of the protein encoded by the gene. Having twice as much of this protein, as in the normal homozygote, does not alter the phenotype. If the allele is incompletely dominant, this means that one copy of the normal allele does not produce the same trait as the homozygote.

C6. The ratio would be 1 normal : 2 star-eyed individuals.

C8. Types O and AB provide an unambiguous genotype. Type O can only be *ii*, and type AB can only be I^AI^B. It is possible for a couple to produce children with all four blood types. The couple would have to be I^Ai and I^Bi. Construct a Punnett square, and you will see that they can produce children with AB, A, B, and O blood types.

C10. A. 1/4 B. 0 C. (1/4)(1/4)(1/4) = 1/64

D. Use the binomial expansion equation, where $n = 3$, $x = 2$, $p = 1/4$, $q = 1/4$. The answer is 0.047, or 4.7%.

$$P = \frac{n!}{x!(n-x)!}p^x q^{(n-x)}$$

$$n = 3,\ p = 1/4,\ q = 1/4,\ x = 2$$

$$P = 3/64 = 0.047, \text{ or } 4.7\%$$

C12. All of the F_1 generation will be white because they have inherited the dominant white allele from their Leghorn parent. Construct a Punnett square. Let *W* and *w* represent one gene, where *W* is dominant and causes a white phenotype. Let *A* and *a* represent the second gene, where the recessive allele causes a white phenotype in the homozygous condition. The genotype of the F_1 birds is *WwAa*. The phenotypic ratio of the F_2 generation will be 13 white : 3 brown. The only brown birds will be 2 *wwAa*, 1 *wwAA*.

C14. We know that the parents must be heterozygotes for both genes.

The genotypic ratio of their children is 1 *BB* : 2 *Bb* : 1 *bb*

The phenotypic ratio depends on sex: 1 *BB* bald male : 1 *BB* bald female : 2 *Bb* bald males : 2 *Bb* nonbald females : 1 *bb* nonbald male : 1 *bb* nonbald female

A. 50% B. 1/8, or 12.5%

C. (3/8)(3/8)(3/8)= 27/512 = 0.05, or 5%

C16. You would look at the pattern within families over the course of many generations. For a recessive trait, 25% of the offspring within a family are expected to be affected if both parents are unaffected carriers, and 50% of the offspring would be affected if one parent was affected. You could look at many families and see if these 25% and 50% values are approximately true. Incomplete penetrance would not necessarily predict such numbers. Also, for very rare alleles, incomplete penetrance would probably have a much higher frequency of affected parents producing affected offspring. For rare recessive disorders, it is most likely that both parents are heterozygous carriers. Finally, the most informative pedigrees would be situations in which two affected parents produce children. If they can produce an unaffected offspring, this would indicate incomplete penetrance. If all of their offspring were affected, this would be consistent with recessive inheritance.

C18. The probability of a heterozygote passing the allele to his or her offspring is 50%. The probability of an affected offspring expressing the trait is 80%. We use the product rule to determine the likelihood of these two independent events.

$$(0.5)(0.8) = 0.4, \text{ or } 40\% \text{ of the time}$$

C20. This is an example of incomplete dominance. The heterozygous horses are palominos. For example, if *C* represents chestnut and *c* represents cremello, the chestnut horses are *CC*, the cremello horses are *cc*, and the palominos are *Cc*.

Application and Experimental Questions

E2. Mexican hairless dogs are heterozygous for a dominant allele that is lethal when homozygous. In a cross between two Mexican hairless dogs, we expect 1/4 to be normal, 1/2 to be hairless, and 1/4 to die before birth.

E4. The first offspring must be homozygous for the horned allele. The father's genotype is still ambiguous; he could be heterozygous or homozygous for the horned allele. The mother's genotype must be heterozygous because her phenotype is polled (she cannot be homozygous for the horned allele) but she produced a horned daughter (who must have inherited a horned allele from its mother).

E6. It is a sex-limited trait, where *W* (white) is dominant but expressed only in females. In this cross of two yellow butterflies, the male is *Ww* but is still yellow because the white phenotype is limited to females. The female is *ww* and yellow. The offspring would be 50% *Ww* and 50% *ww*. All males would be yellow. Half of the females would be white (*Ww*) and half would be yellow (*ww*). Overall, this would yield 50% yellow males, 25% yellow females, and 25% white females.

E8. The yellow squash has to be *wwgg*. It has to be *ww* because it is colored, and it has to be *gg* because it is yellow. Because the cross produced 50% white and 50% green offspring, the other parent (i.e., the white squash) must be *WwGG*. The cross would produce 50% offspring that are *WwGg* (white) and 50% that are *wwGg* (green).

E10. In this cross, we expect a 9:7 ratio between red and white flowers. In other words, 9/16 will be red and 7/16 will be white. Because there are a total of 345 plants, the expected values are

$$9/16 \times 345 = 194 \text{ red}$$
$$7/16 \times 345 = 151 \text{ white}$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{(201 - 194)^2}{194} + \frac{(144 - 151)^2}{151}$$

$$\chi^2 = 0.58$$

With 1 degree of freedom, our chi square value is too small to reject our hypothesis. Therefore, we accept that it may be correct.

Questions for Student Discussion/Collaboration

2. Let's refer to the alleles as *B* dominant, *b* recessive and *G* dominant, *g* recessive.

The parental cross is *BBGG* × *bbgg*.

All of the F_1 are *BbGg*.

Make a Punnett square for *BbGg* × *BbGg*

According to your Punnett square, the genotypes that are homozygous for the *b* allele and have at least one copy of the dominant *G* allele are gray. To explain this phenotype, we could hypothesize that the *B* allele encodes an enzyme that can make lots of pigment, whether or not the *G* allele is present. Therefore, you get a black phenotype when one *B* allele is inherited. The *G* allele encodes a somewhat redundant enzyme, but maybe it does not function quite as well or its pigment product may not be as dark. Therefore, in the absence of a *B* allele, the *G* allele will give a gray pigment.

CHAPTER 6

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 6.1 A nucleoid is not surrounded by a membrane as is the cell nucleus.

FIGURE 6.2 The chloroplast genome contains about 100 genes, whereas nuclear genomes in plants have several thousand or tens of thousands of genes.

FIGURE 6.3 A reciprocal cross is a cross in which the sexes and phenotypes of the parents are reversed relative to a first cross.

FIGURE 6.4 No. Once a patch of tissue is white, it has lost all of the normal chloroplasts, so it could not produce a patch of green tissue.

FIGURE 6.5 The mt^+ mating type transmits chloroplasts to the offspring.

FIGURE 6.6 One mitochondrial nucleoid usually contains multiple copies of the chloroplast genome.

FIGURE 6.7 Mitochondria need rRNA and tRNA genes to translate proteins within the mitochondrial matrix.

FIGURE 6.8 Chloroplast and mitochondrial genomes have lost most of their genes during evolution. Many of these have been transferred to the cell nucleus.

FIGURE 6.9 All of the offspring would be normal because they would inherit an active copy from their father.

FIGURE 6.10 Erasure allows eggs to transmit unmethylated copies of the gene to the offspring.

FIGURE 6.11. Methylation prevents CTC-binding factor from binding to the ICR.

FIGURE 6.12. Maintenance methylation is automatic methylation that occurs when a methylated gene replicates and is transferred to daughter cells. It occurs in somatic cells. *De novo* methylation is the methylation of a gene that is not already methylated. It occurs in germline cells.

FIGURE 6.13 The offspring on the left did not receive a copy of either gene from their mother. Their father silences the AS gene, so they have AS because they do not have an active copy of the AS gene. By comparison, they receive an active PWS gene from their father, which prevents them from having PWS.

FIGURE 6.14 The offspring are all dextral because all of the F_1 mothers are *Dd*, and the genotype of the mother determines the phenotype of the offspring.

FIGURE 6.15 The egg cell will receive both *D* and *d* gene products.

End-of-Chapter Questions:

Conceptual Questions

C2. Extranuclear inheritance does not always occur via the female gamete. Sometimes it occurs via the male gamete. Even in species in which maternal inheritance is prevalent, paternal leakage can also occur, in which the paternal parent occasionally provides mitochondria via the sperm. Maternal inheritance is the most common form of extranuclear inheritance because the female gamete is relatively large and more likely to contain cell organelles.

C4. The mitochondrial and chloroplast genomes are composed of a circular chromosome found in one or more copies in a region of the organelle known as the nucleoid. The number of genes per chromosome varies from species to species. Chloroplasts genomes tend to be larger than mitochondria genomes. See Tables 6.2 and 6.3 for examples of the variation among mitochondrial and chloroplast genomes.

C6. A. Yes. B. Yes.

C. No, it is determined by a gene in the chloroplast genome.

D. No, it is determined by a gene in the mitochondria.

C8. Biparental extranuclear inheritance would resemble Mendelian inheritance in that offspring could inherit alleles of a given gene from both parents. It differs, however, when you think about it from the perspective of heterozygotes. For a Mendelian trait, the law of segregation tells us that a heterozygote passes one allele for a given gene to an offspring, but not both. In contrast, if a parent has a mixed population of mitochondria (e.g., some carrying a mutant gene and some carrying a normal gene), that parent could pass both types of genes (mutant and normal) to a single offspring, because more than one mitochondrion could be contained within a sperm or egg cell.

C10. Genomic imprinting is a pattern in which a modification occurs to a gene or chromosome that alters gene expression, but the pattern is different during oogenesis versus spermatogenesis. This leads to monoallelic expression. Examples include the *Igf2* gene, and genes associated with Prader-Willi and Angelman syndromes.

C12. The *de novo* methylation would not occur in somatic cells but may occur in germ cells. Somehow, the methylase must be able to recognize particular DMRs, because we know that some genes are imprinted in males but not females, whereas other genes are imprinted in females but not males.

C14. In some species, such as marsupials, X inactivation depends on the sex. This is similar to imprinting. Also, once X inactivation occurs during embryonic development, it is remembered throughout the rest of the life of the organism, which is also similar to imprinting. X inactivation in mammals is different from genomic imprinting, in that it is not sex-dependent. The X chromosome that is inactivated could be inherited from the mother or the father. There was no marking process on the X chromosome that occurred during gametogenesis. In contrast, genomic imprinting always involves a marking process during gametogenesis.

C16. A. Genotypes: All *Nn*

Phenotypes: All abnormal

B. Genotypes: All *Nn*

Phenotypes: All normal

C. Genotypes: 1 *NN* : 2 *Nn* : 1 *nn*

Phenotypes: All normal (because the mother is heterozygous and *N* is dominant)

C18. A. For an imprinted gene, you need to know whether the mutant *Igf2* allele is inherited from the mother or the father.

B. For a maternal-effect gene, you need to know the genotype of the mother.

C. For maternal inheritance, you need to know whether the mother has this trait, because the offspring will inherit her mitochondria.

C20. The mother is *hh*. We know this because it is a maternal-effect gene and all of its offspring have small heads. The offspring are all *hh* because their mother is *hh* and their father is *hh*.

Application and Experimental Questions

E2. In the absence of UV light, we would expect all sm^r offspring. With UV light, we would expect a greater percentage of sm^s offspring.

E4. The results of each succeeding generation depended on the genotypes of the mothers of the preceding generation. For example, if a mother was *dd*, the F_1 offspring were all sinistral. The genotypes of the F_2 mothers were 1 *DD* : 2 *Dd* : 1 *dd*. The *DD* and *Dd* mothers produced dextral offspring of the F_3 generation, and the *dd* mothers produced sinistral offspring. As expected, the ratio of dextral to sinistral was 3:1, which is derived from the genotypes of the F_2 mothers, which were 1 *DD* : 2 *Dd* : 1 *dd*.

E6. The pattern of inheritance is consistent with imprinting. In every cross, the allele that is inherited from the father is expressed in the offspring, whereas the allele inherited from the mother is not.

Questions for Student Discussion/Collaboration

2. Obviously, you cannot maintain a population of flies that are homozygous for a recessive lethal allele. However, heterozygous females can produce viable offspring and these can be crossed to heterozygous or homozygous males to produce homozygous females. An experimenter would routinely have to make crosses and determine that the recessive allele was present in a population of flies by identifying homozygous females that were unable to produce any viable offspring. Maintaining a population of flies that carry a lethal recessive allele can be much easier if the recessive lethal allele is closely linked to a dominant allele that is not lethal, such as one affecting eye color or some other trait. If a fly exhibits the dominant trait, it is likely that it is also carrying the recessive lethal allele.

CHAPTER 7

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 7.1 The offspring found in excess are those with purple flowers, long pollen, and red flowers, round pollen.

FIGURE 7.2 No, such a crossover would not change the arrangements of these alleles.

FIGURE 7.3 A single crossover can produce offspring with: gray body, red eyes, miniature wings; gray body, white eyes, miniature wings; yellow body, red eyes, long wings; and yellow body, white eyes, long wings.

FIGURE 7.4 When genes are relatively close together, a crossover is relatively unlikely to occur between them. Therefore, the nonrecombinant offspring are more common.

FIGURE 7.5 The reason is because the *w* and *m* genes are farther apart than the *y* and *w* genes.

FIGURE 7.6 Crossing over can change the combination of kernel phenotypes, and also it can change the morphologies of the chromosomes compared with those of the nonrecombinant chromosomes.

FIGURE 7.7 Genetic mapping is useful for (1) understanding the complexity and genetic organization of a species; (2) understanding the underlying basis of inherited traits; (3) cloning genes; (4) understanding evolution; (5) diagnosing and treating diseases: (6) predicting the likelihood of a couple having offspring with genetic diseases; (7) breeding agricultural livestock and crops.

FIGURE 7.8 Crossing over occurred during oogenesis in the female parent of the recombinant offspring.

FIGURE 7.9 The basic strategy is to conduct testcrosses, like the one shown in Figure 7.8. The percentage of recombinant offspring is a measure of the map distance between the genes.

FIGURE 7.10 Multiple crossovers prevent the maximum percentage of recombinant offspring from exceeding 50%.

FIGURE 7.11 Mitotic recombination occurs in somatic cells.

End-of-Chapter Questions:

Conceptual Questions

C2. An independent assortment hypothesis is used because it enables us to calculate the expected values based on Mendel's ratios. Using the observed and expected values, we can calculate whether or not the deviations between the observed and expected values are too large to occur as a matter of chance. If the deviations are very large, we reject the hypothesis of independent assortment.

C4. A single crossover produces *A B C*, *A b c*, *a B C*, and *a b c*.

A. Between 2 and 3, between genes *B* and *C*

B. Between 1 and 4, between genes *A* and *B*

C. Between 1 and 4, between genes *B* and *C*

D. Between 2 and 3, between genes *A* and *B*

C6. The likelihood of scoring a basket would be greater if the basket was larger. Similarly, the chances of a crossover initiating in a region between two genes is proportional to the size of the region between the two genes. There are a finite number (usually a few) of crossovers that occur between homologous chromosomes during meiosis, and the likelihood that a crossover will occur in a region between two genes depends on how big that region is.

C8. The pedigree suggests a linkage between the dominant allele causing nail-patella syndrome and the I^B allele of the ABO blood type gene. In every case, the individual who inherits the I^B allele also inherits this disorder.

C10. Ass–1 43 Sdh–1 5 Hdc–9 Hao–1 6 Odc–2 8 Ada–1

C12. The inability to detect double crossovers causes the map distance to be underestimated. In other words, more crossovers occur in the region than we realize. When we have a double crossover, we do not get a recombinant offspring (in a dihybrid cross). Therefore, the second crossover cancels out the effects of the first crossover.

C14. Mitotic recombination is crossing over between homologous chromosomes during mitosis in somatic cells. Mitotic recombination is one explanation for a blue patch. Following mitotic recombination, the two chromosomes carrying the *b* allele could segregate into the same cell and produce the blue color. Another reason could be chromosome loss; the chromosome carrying the *B* allele could be lost during mitosis.

Application and Experimental Questions (Includes Most Mapping Questions)

E2. They could have used a strain with two abnormal chromosomes. In this case, the recombinant chromosomes would either look normal or have abnormalities at both ends.

E4. The rationale behind a testcross is to determine if recombination has occurred during meiosis in the heterozygous parent. The other parent is usually homozygous recessive, so we cannot tell if crossing over has occurred in the recessive parent. It is easier to interpret the data if a testcross does use a completely homozygous recessive parent. However, in the other parent, it is not necessary for all of the dominant alleles to be on one chromosome and all of the recessive alleles on the other. The parental generation provides us with information concerning the original linkage pattern between the dominant and recessive alleles.

E6. The reason why the percentage of recombinant offspring is more accurate when the genes are close together is because fewer double crossovers occur. The inability to detect double crossover causes the map distance to be underestimated. If two genes are very close together, very few double crossovers occur, so that underestimation due to double crossovers is minimized.

E8. Morgan determined this by analyzing the data in gene pairs. This analysis revealed that there were fewer recombinants between certain gene pairs (e.g., body color and eye color) than between other gene pairs (e.g., eye color and wing length). From this comparison, he hypothesized that genes that are close together on the same chromosome will produce fewer recombinants than genes that are farther apart.

E10. Map distance:

$$= \frac{64 + 58}{333 + 64 + 58 + 380} \times 100$$

$$= 15.1 \text{ mu}$$

E12. We consider the genes in pairs: there should be 10% offspring due to crossing over between genes *A* and *B*, and 5% due to crossing over between *A* and *C*.

A. This is due to a crossover between *B* and *A*. The nonrecombinants are *Aa bb Cc* and *aa Bb cc*. The 10% recombinants are *Aa Bb Cc* and *aa bb cc*. If we assume an equal number of both types of recombinants, 5% are *Aa Bb Cc*.

B. This is due to a crossover between *A* and *C*. The nonrecombinants are *Aa bb Cc* and *aa Bb cc*. The 5% recombinants are *aa Bb Cc* and *Aa bb cc*. If we assume an equal number of both types of recombinants, 2.5% are *aa Bb Cc*.

C. This is also due to a crossover between *A* and *C*. The nonrecombinants are *Aa bb Cc* and *aa Bb cc*. The 5% recombinants are *aa Bb Cc* and *Aa bb cc*. If we assume an equal number of both types of recombinants, 2.5% are *Aa bb cc*.

E14. A. One basic strategy to solve this problem is to divide the data up into gene pairs and determine the map distance between two genes.

1. 184 tall, smooth	2. 153 tall, normal	3. 163 smooth, normal
13 tall, peach	44 tall, oblate	33 smooth, oblate
184 dwarf, peach	155 dwarf, oblate	31 peach, normal
12 dwarf, smooth	41 dwarf, normal	166 peach, oblate

$$1.\ \text{Map distance} = \frac{13 + 12}{184 + 13 + 184 + 12} \times 6.4 \text{ mu}$$

$$2.\ \text{Map distance} = \frac{44 + 41}{153 + 44 + 155 + 41} \times 21.6 \text{ mu}$$

$$3.\ \text{Map distance} = \frac{33 + 31}{163 + 33 + 31 + 166} \times 16.3 \text{ mu}$$

Use the two shortest distances to compute the map:

Tall, dwarf 6.4 Smooth, peach 16.3 Normal, oblate

E16. In the backcross, the two nonrecombinant types would be the homozygotes that cannot make either enzyme and the heterozygotes that can make both enzymes. The recombinants would make one enzyme but not both. Because the two genes are 12 mu apart, 12% would be recombinants and 88% would be nonrecombinant types. Because there are two nonrecombinant types produced in equal numbers, we would expect 44% of the mice to be unable to make either enzyme.

E18. A. If we represent *B* (bushy tail) and *b* (normal tail) for one gene, and *Y* (yellow) and *y* (white) for the second gene:

Parent generation: *BBYY* × *bbyy*

F_1 generation: All *BbYy* (Note: if the two genes are linked, *B* would be linked to *Y* and *b* would be linked to *y*.)

Testcross: F_1 *BbYy* × *bbyy*

Nonrecombinant offspring from testcross: *BbYy* and *bbyy*

BbYy males—bushy tails, yellow

bbyy males—normal tails, white

BbYy females—normal tails, yellow

bbyy females—normal tails, white

Recombinant offspring from testcross: *Bbyy* and *bbYy*

Bbyy males—bushy tails, white

bbYy males—normal tails, yellow

Bbyy females—normal tails, white

bbYy females—normal tails, yellow

We cannot use the data regarding female offspring, because we cannot tell if females are recombinant or nonrecombinant, because all females have normal tails. However, we can tell if male offspring are recombinant.

If we use the data on males to conduct a chi square analysis, we expect a 1:1:1:1 ratio among the male offspring. Because there are 197 male offspring total, we expect 1/4, or 49 (rounded to the nearest whole number), of the four possible phenotypes. To compute the chi square:

$$\chi^2 = \frac{(28-49)^2}{49} + \frac{(72-49)^2}{49} + \frac{(68-49)^2}{49} + \frac{(29-49)^2}{49}$$

$$\chi^2 = 9.0 + 10.8 + 7.4 + 8.2$$

$$\chi^2 = 35.4$$

If we look up the value of 35.4 in our chi square table, with 3 degrees of freedom, the value lies far beyond the 0.01 probability level. Therefore, it is very unlikely to get such a large deviation if our hypothesis of independent assortment is correct. Therefore, we reject our hypothesis and conclude that the genes are linked.

B. To compute map distance:

$$\frac{28 + 29}{28 + 72 + 68 + 29} \times 100 = 28.9 \text{ mu}$$

E20. Let's use the following symbols: *S* for normal nose, *s* for snubnose, *p* for normal tail, *P* for pintail, *J* for normal gait, *j* for jerker.

The parental cross is *ss PP jj* crossed to *SS pp JJ*.

The F_1 offspring would all be *Ss Pp Jj*. If the genes are linked, the alleles *s*, *P*, and *j* would be linked on one chromosome, and the alleles *S*, *p*, and *J* would be linked on the homologous chromosome.

The testcross is F_1 mice, which are *Ss Pp Jj*, crossed to *ss pp jj* mice.

To measure the distances between the genes, we can separate the data into gene pairs.

Nose shape, tail length

631 snubnose, pintail—nonrecombinant

111 snubnose, normal tail—recombinant

625 normal nose, normal tail—nonrecombinant

115 normal nose, pintail—recombinant

$$\text{Map distance} = \frac{111 + 115}{631 + 111 + 625 + 115} \times 100 = 15.2 \text{ mu}$$

Nose shape, normal gait

662 snubnose, jerker—nonrecombinant

80 snubnose, normal gait—recombinant

652 normal nose, normal gait—nonrecombinant

88 normal nose, jerker—recombinant

$$\text{Map distance} = \frac{80 + 88}{662 + 80 + 652 + 88} \times 100 = 11.3 \text{ mu}$$

Tail length, normal gait

571 pintail, jerker—nonrecombinant

175 pintail, normal gait—recombinant

557 normal tail, normal gait—nonrecombinant

179 normal tail, jerker—recombinant

$$\text{Map distance} = \frac{175 + 179}{571 + 175 + 557 + 179} \times 100 = 23.9 \text{ mu}$$

The order of the genes is tail length, nose shape, and normal gait (or you could say the opposite order). Nose shape is in the middle.

If we use the two shortest distances to construct our map:

$\underline{P \quad 15.2 \quad S \quad 11.3 \quad J}$

E22. A.

Parent | Parent
$\underline{b\ 7\ A\ 4\ C}$ | $\underline{B\ 7\ a\ 4\ c}$
$\underline{b\ 7\ A\ 4\ C}$ | $\underline{B\ 7\ a\ 4\ c}$

↓

Offspring
$\underline{b\ 7\ A\ 4\ C}$
$\underline{B\ 7\ a\ 4\ c}$

B. A heterozygous F_2 offspring would have to inherit a chromosome carrying all of the dominant alleles. In the F_1 parent (of the F_2 offspring), a crossover in the 7 mu region between genes *b* and *A* (and between *B* and *a*) would yield a chromosome that was *B A C* and *b a c*. If an F_2 offspring inherited the *B A C* chromosome from its F_1 parent and the *b a c* chromosome from the homozygous parent, it would be heterozygous for all three genes.

C. If you look at the answer to part B, a crossover between genes *b* and *A* (and between *B* and *a*) would yield *B A C* and *b a c* chromosomes. If an offspring inherited the *b a c* chromosome from its F_1 parent and the *b a c* chromosome from its homozygous parent, it would be homozygous for all three genes. The chances of a crossover in this region are 7%. However, half of this 7% crossover event yields chromosomes that are *B A C*, and the other half yields chromosomes that are *bac*. Therefore, the chances are 3.5% of getting homozygous F_2 offspring.

Questions for Student Discussion/Collaboration

2. The X and Y chromosomes are not completely distinct linkage groups. One might describe them as overlapping linkage groups having some genes in common, but most genes which are not common to both.

CHAPTER 8

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 8.1 The staining of chromosomes results in banding patterns that make it easier to distinguish chromosomes that are similar in size and have similar centromere locations.

FIGURE 8.2 Deletions and duplications alter the total amount of genetic material.

FIGURE 8.3 If a chromosomal fragment does not contain a centromere, it will not segregate properly. If it remains outside the nucleus, it will be degraded.

FIGURE 8.5 Nonallelic homologous recombination occurs because of the pairing of homologous sites that duplicated within the chromosomes. This pairing causes the chromosomes to be misaligned.

FIGURE 8.10 These chromosomes form an inversion loop because the homologous regions are pairing with each other. For the inverted and noninverted regions to pair, a loop must form.

FIGURE 8.11 The mechanism shown in part (b) may occur if transposable elements are found in different chromosomes. These elements may promote the pairing between nonhomologous chromosomes and a subsequent crossover could occur.

FIGURE 8.13 Two out of six gametes (the two on the left) will produce a viable offspring with a normal phenotype. Therefore, the probability is 2/6, or 1/3.

FIGURE 8.14 These chromosomes form a translocation cross because homologous regions are pairing with each other.

FIGURE 8.15 Aneuploid, monosomic, monosomy 3.

FIGURE 8.16 The genes on chromosome 2 would be found in single copies, whereas the genes on the other chromosomes would be found in two copies. The expression of genes on chromosome 2 would be less (perhaps 50%) relative to a normal individual. This creates an imbalance between genes on chromosome 2 and those on the other chromosomes.

FIGURE 8.19 About 512.

FIGURE 8.20 Polyploid plants are often more vigorous and disease-resistant. They may have larger flowers and produce more fruit.

FIGURE 8.21 During meiosis in a triploid individual, the homologs cannot pair properly. This results in highly aneuploid gametes, which are usually nonviable. Also, if aneuploid gametes participate in fertilization, the offspring are usually nonviable.

FIGURE 8.22 In this case, nondisjunction means that pairs of chromosomes are not separating from each other properly.

FIGURE 8.25 In autopolyploidy, multiple sets of chromosomes come from the same species. In allopolyploidy, multiple sets of chromosomes come from at least two different species.

FIGURE 8.27 The allotetraploid is more likely to produce viable gametes because the chromosomes can pair up and produce euploid gametes. By comparison, the allodiploid will produce aneuploid gametes that are likely to be nonviable.

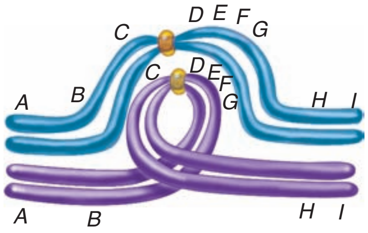
End-of-Chapter Questions:

Conceptual Questions

C2. Small deletions and duplications are less likely to affect phenotype simply because they usually involve fewer genes. If a small deletion did have a phenotypic effect, you would conclude that a gene or genes in this region are required to have a normal phenotype.

C4. A gene family is a group of genes that are derived from the process of gene duplications. They have similar sequences, but the sequences have some differences due to the accumulation of mutations over many generations. The members of a gene family usually encode proteins with similar but specialized functions. The specialization may occur in different cells or at different stages of development.

C6. It has a pericentric inversion.



C8. There are four products from meiosis. One would be a normal chromosome, and one would contain the inversion described in the question of C8. The other two chromosomes would be dicentric or acentric with the following order of genes:

centromere centromere

$\underline{A \quad \downarrow \quad B\,C\,D\;\;E\,F\;\;G\,H\;\;I\,J\;\;\;D\;\;C\;\;B \qquad \downarrow \qquad A}$ Dicentric

$\underline{M\;\;L\;\;K\;\;\;J\,I\;\;H\;\;G\;\;F\;\;\;E\;\;\;K\;\;L\;\;M}$ Acentric

C10. In the absence of crossing over, alternative segregation would yield half of the cells with two normal chromosomes and half with a balanced translocation. For adjacent-1 segregation, all cells will be unbalanced. Two cells would be

$\underline{A\,B\,C\,D\,E} + \underline{A\,I\,J\,K\,L\,M}$

And the other two cells would be:

$\underline{H\,B\,C\,D\,E} + \underline{H\,I\,J\,K\,L\,M}$

C12. One of the parents may carry a balanced translocation between chromosomes 5 and 7. The phenotypically abnormal offspring has inherited an imbalanced translocation due to the segregation of translocated chromosomes during meiosis.

C14. A deficiency and an unbalanced translocation are more likely to have phenotypic effects because they create genetic imbalances. For a deficiency, there are too few copies of several genes, and for an unbalanced translocation, there are too many.

C16. It is due to a crossover within the inverted region. You should draw the inversion loop. The crossover occurred between *P* and *U*.

C18. This person has a total of 46 chromosomes. However, this person would be considered aneuploid rather than euploid, because one of the sets is missing a sex chromosome and one set has an extra copy of chromosome 21.

C20. It may be related to genetic balance. In aneuploidy, there is an imbalance in gene expression between the chromosomes found in their normal copy number versus those that are either too many or too few. In polyploidy, the balance in gene expression is still maintained.

C22. Offspring with trisomies 13, 18, and 21 survive because the chromosomes are small and contain fewer genes than the larger chromosomes. Individuals with abnormal numbers of X chromosomes can survive because the extra copies are converted to transcriptionally inactive Barr bodies. The other trisomies are lethal because they cause a great amount of imbalance between the level of gene expression on the normal diploid chromosomes relative to the chromosomes that are trisomic.

C24. Endopolyploidy means that a particular somatic tissue is polyploid even though the rest of the organism is not. The biological significance is not entirely understood, although it has been speculated that an increase in chromosome number in certain cells may enhance their ability to produce specific gene products that are needed in great abundance.

C26. In certain types of cells, such as salivary cells, the homologous chromosomes pair with each other and then replicate approximately nine times to produce a polytene chromosome. The centromeres from each chromosome aggregate with each other at the chromocenter. This structure has six arms that arise from the single arm of two telocentric chromosomes (the X and 4) and two arms each from the metacentric chromosomes 2 and 3.

C28. Aneuploid should not be used.

C30. There are 11 chromosomes per set, so there are 11 possible trisomies: trisomy 1, trisomy 2, trisomy 3, trisomy 4, trisomy 5, trisomy 6, trisomy 7, trisomy 8, trisomy 9, trisomy 10, and trisomy 11.

C32. The odds of producing a euploid gamete are $(1/2)^{n-1}$, which equals $(1/2)^5$, or a 1 in 32 chance. The chance of producing an aneuploid gamete is 31 out of 32 gametes. We use the product rule to determine the chances of getting a euploid individual because a euploid individual is produced from two euploid gametes: $1/32 \times 1/32 = 1/1024$. In other words, if this plant self-fertilized, only 1 in 1024 offspring would be euploid. The euploid offspring could be diploid, triploid, or tetraploid.

C34. It usually occurs during meiosis I when the homologs synapse to form bivalents. During meiosis II and mitosis, the homologs do not synapse. Instead, the chromosomes align randomly along the metaphase plate and then the centromeres separate.

C36. An allodiploid is an organism having one set of chromosomes from two different species. Unless the two species are closely related evolutionarily, the chromosomes do not synapse during meiosis. Therefore, they do not segregate properly. This produces aneuploid gametes that are usually inviable. By comparison, allotetraploids that have two sets of chromosomes from each species are more likely to be fertile because each chromosome has a homolog to pair with during meiosis.

C38. For meiotic nondisjunction, the bivalents are not separating correctly during meiosis I. During mitotic nondisjunction, the sister chromatids are not separating properly.

C40. Complete nondisjunction occurs during meiosis I so that one nucleus receives all the chromosomes and the other nucleus does not get any. The nucleus with all the chromosomes then proceeds through a normal meiosis II to produce two haploid sperm cells.

Experimental Questions

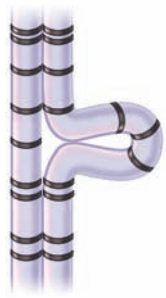
E2. Colchicine interferes with the spindle apparatus, thereby causing nondisjunction. At high concentrations, it can cause complete nondisjunction and produce polyploid cells.

E4. A. The F_1 offspring would probably be phenotypically normal, because they would carry the correct number of genes.

B. The F_1 offspring would have lowered fertility because they are inversion heterozygotes. Because this is a large inversion, crossing over is fairly likely in the inverted region. When this occurs, it will produce deletions and duplications that will probably be lethal in the resulting F_2 offspring.

E6. First, you would cross the two strains together. It is difficult to predict the phenotype of the offspring. Nevertheless, you would keep crossing offspring to each other and backcrossing them to the parental strains until you obtained a great-tasting tomato strain that was resistant to heat and the viral pathogen. You could then make this strain tetraploid by treatment with colchicine. If you crossed the tetraploid strain with your great-tasting diploid strain that was resistant to heat and the viral pathogen, you may get a triploid that had these characteristics. This triploid would probably be seedless.

E8. A polytene chromosome is formed when a chromosome replicates many times, and the chromatids lie side by side, as shown in Figure 8.18. The homologous chromosomes also lie side by side. Therefore, if there is a deletion, there will be a loop. The loop is the segment that is not deleted from one of the two homologs.



Questions for Student Discussion/Collaboration

2. There are lots of possibilities. Students could look in agriculture and botany books to find many examples. In the insect world, there are interesting examples of euploidy affecting gender determination. Among amphibians and reptiles, there are also several examples of closely related species that have euploid variation.

4. 1. Polyploids are often more robust and disease-resistant.

   2. Allopolyploids may have useful combinations of traits.

   3. Hybrids are often more vigorous; they can be generated from monoploids.

   4. Strains with an odd number of chromosome sets (e.g., triploids) are usually seedless.

CHAPTER 9

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 9.1 To grow, the colonies must have functional copies of all five genes. This could occur by the transfer of the met^+ and bio^+ genes to the $met^-bio^-thr^+leu^+thi^+$ strain or the transfer of the thr^+, leu^+, and thi^+ genes to the $met^+bio^+thr^-leu^-thi^-$ strain.

FIGURE 9.2 Because bacteria are too small to pass through the filter, the apparatus can determine if direct cell-to-cell contact is necessary for gene transfer to occur.

FIGURE 9.3 It would be found in an F^+ cell.

FIGURE 9.4 Relaxase is a part of the relaxosome, which is needed for the cutting of the F factor and its transfer to the recipient cell. The coupling factor guides the DNA strand to the exporter, which transports it to the recipient cell.

FIGURE 9.5 An F' factor carries a portion of the bacterial chromosome, whereas an F factor does not.

FIGURE 9.6 Because conjugation occurred for a longer period of time, pro^+ was transferred in the conjugation experiment shown in the bottom right.

FIGURE 9.7 Interrupted mating was achieved by blender treatment. This was done at different time points.

FIGURE 9.8 This type of map is based on the time of gene transfer in conjugation experiments, which is determined in minutes.

FIGURE 9.9 The *lacZ* gene is closer to the origin of transfer; its transfer began at 16 minutes.

FIGURE 9.10 The normal process is for bacteriophage DNA to be incorporated into a phage coat. In transduction, a segment of bacterial chromosomal DNA is incorporated into a phage coat.

FIGURE 9.11 If the two genes were very far apart, the arg^+ gene would never be cotransduced with the met^+ gene.

FIGURE 9.12 If the recipient cell did not have a lys^- gene, the lys^+ gene could be incorporated by nonhomologous recombination.

End-of-Chapter Questions:

Conceptual Questions

C2. It is not a form of sexual reproduction, because two distinct parents do not produce gametes that unite to form a new individual. However, conjugation is similar to sexual reproduction in the sense that the genetic material from two cells are somewhat mixed. In conjugation, there is not the mixing of two genomes, one from each gamete. Instead, there is a transfer of genetic material from one cell to another. This transfer can alter the combination of genetic traits in the recipient cell.

C4. An F^+ strain contains a separate, circular piece of DNA that has its own origin of transfer. An *Hfr* strain has its origin of transfer integrated into the bacterial chromosome. An F^+ strain can transfer only the DNA contained on the F factor. If given enough time, an *Hfr* strain can actually transfer the entire bacterial chromosome to the recipient cell.

C6. Sex pili promote the binding of donor and recipient cells.

C8. Cotransduction is the transduction of two or more genes. The distance between the genes determines the frequency of cotransduction. When two genes are close together, the cotransduction frequency would be higher than that for two genes that are relatively farther apart.

C10. If a site that frequently incurred a breakpoint was between two genes, the cotransduction frequency of these two genes would be much less than expected. This is because the site where the breakage occurred would separate the two genes from each other.

C12. The transfer of conjugative plasmids such as F factor DNA.

C14. A. Transformation is the most likely mechanism because conjugation does not usually occur between different species, particularly distantly related species, and different species are not usually infected by the same bacteriophages.

B. It could occur in a single step, but it may be more likely to have involved multiple steps.

C. The use of antibiotics selects for the survival of bacteria that have resistance genes. If a population of bacteria is exposed to an antibiotic, those carrying resistance genes will survive and their relative numbers will increase in subsequent generations.

Experimental Questions

E2. Mix the two strains together and then put some of them on plates containing streptomycin and some of them on plates without streptomycin. If mated colonies are present on both types of plates, then the *thr*$^+$, *leu*$^+$, and *thi*$^+$ genes were transferred to the *met*$^+$ *bio*$^+$ *thr*$^-$ *leu*$^-$ *thi*$^-$ strain. If colonies are found only on the plates that lack streptomycin, then the *met*$^+$ and *bio*$^+$ genes are being transferred to the *met*$^-$ *bio*$^-$ *thr*$^+$ *leu*$^+$ *thi*$^+$ strain. This answer assumes a one-way transfer of genes from a donor to a recipient strain.

E4. An interrupted mating experiment is a procedure in which two bacterial strains are allowed to mate, and then the mating is interrupted at various time points. The interruption occurs by agitation of the solution in which the bacteria are found. This type of study is used to map the locations of genes. It is necessary to interrupt mating so that you can vary the time and obtain information about the order of transfer; which gene transferred first, second, and so on.

E6. Mate unknown strains *A* and *B* to the F^- strain in your lab that is resistant to streptomycin and cannot use lactose. This is done in two separate tubes (i.e., strain *A* plus your F^- strain in one tube, and strain *B* plus your F^- strain in the other tube). Plate the mated cells on growth media containing lactose plates plus streptomycin. If you get growth of colonies, the unknown strain had to be strain *A*, the F^+ strain that had lactose utilization genes on its F factor.

E8. A. If we extrapolate these lines back to the *x*-axis, the *hisE* intersects at about 3 minutes and the *pheA* intersects at about 24 minutes. These are the values for the times of entry. Therefore, the distance between these two genes is 21 minutes (i.e., 24 minus 3).

B. *hisE* *pabB* *pheA*

4 17

E10. One possibility is that you could treat the P1 lysate with DNase I, an enzyme that digests DNA. (Note: If DNA were digested with DNase I, the function of any genes within the DNA would be destroyed.) If the DNA were within a P1 phage, it would be protected from DNase I digestion. This would allow you to distinguish between transformation (which would be inhibited by DNase I) versus transduction (which would not be inhibited by DNase I). Another possibility is that you could try to fractionate the P1 lysate. Naked DNA would be smaller than a P1 phage carrying DNA. You could try to filter the lysate to remove naked DNA, or you could subject the lysate to centrifugation and remove the lighter fractions that contain naked DNA.

E12. Cotransduction frequency = $(1 - d/L)^3$

For the normal strain:

Cotransduction frequency = $(1 - 0.7/2)^3 = 0.275$, or 27.5%

For the new strain:

Cotransduction frequency = $(1 - 0.7/5)^3 = 0.64$, or 64%

The experimental advantage is that you could map genes that are farther than 2 minutes apart. You could map genes that are up to 5 minutes apart.

E14. Cotransduction frequency = $(1 - d/L)^3$

E16. A. We first need to calculate the cotransformation frequency, which equals 2/70, or 0.029.

B. As you may have expected, the cotransformation frequency is much higher when the transformation involves larger pieces of DNA.

Questions for Student Discussion/Collaboration

2. Conjugation requires direct cell contact, which is mediated by proteins that are found in the same species. So, it is less likely to occur between different species unless they are very closely related evolutionarily. Similarly, gene transfer via transduction involves bacteriophages that are usually species-specific. Gene transfer via transformation is most likely to occur between different species. Gene transfer has several consequences, including antibiotic resistance and the ability to survive under new growth conditions.

CHAPTER 10

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 10.1 One way that viruses vary is in their genomes. Viral genomes vary with regard to size, RNA versus DNA, and single-stranded versus double-stranded. The structures of viruses also vary with regard to the complexity of their capsids and whether or not they contain an envelope.

FIGURE 10.2 A virus may remain latent following its integration into the genome of the host cell.

FIGURE 10.3 The lytic cycle produces new phage particles.

FIGURE 10.6 The genetic material that determines the structure of the virus is found inside the virus itself.

FIGURE 10.7 As more viruses are made, they are released from host cells and eventually cause neighboring cells to lyse. This makes the plaque grow larger.

FIGURE 10.9 Noncomplementation indicates that two mutations are in different genes, whereas complementation indicates that the two mutations are in the same gene.

FIGURE 10.11 The mutant phages that infected E. coli K12(λ) would not be able to propagate and form plaques unless intragenic recombination produced a wild-type phage. This intragenic recombination would be a very rare event, which is why more plaques were used to infect this strain.

FIGURE 10.12 Deletion mapping makes it unnecessary to make all of the possible pairwise crosses to map the locations of the mutations.

FIGURE 10.13 This figure is illustrating the locations of mutations within the *rIIA* or *rIIB* gene.

End-of-Chapter Questions:

Conceptual Questions

C2. All viruses have a nucleic acid genome and a capsid composed of protein. Some eukaryotic viruses are surrounded by an envelope that is composed of a membrane with embedded proteins.

C4. A viral envelope is composed of a membrane with embedded proteins. It is made when the virus buds from the host cell, taking with it a portion of the host cell's plasma membrane.

C6. The attachment step usually involves the binding of the virus to a specific protein on the surface of the host cell. Only certain cell types will make that specific protein.

C8. Reverse transcriptase is used to copy the viral RNA into DNA so it can be integrated into a chromosome of the host cell.

C10. A temperate phage can follow either the lytic or lysogenic cycle, whereas a virulent phage can follow only the lytic cycle.

C12. The term *complementation* refers to the outcome when two mutations are combined in a single individual. If two defective mutations combined in a single individual result in a nondefective phenotype, complementation has occurred. This is because the two mutations are in different genes, and the corresponding wild-type versions are dominant. If complementation does not occur, this means that the defective mutations are in the same gene. In this case, when combined in the same individual, complementation cannot occur because the individual does not have a wild-type (functional) copy of the gene.

C14. Perhaps the mutations are at exactly the same spot (i.e., the same nucleotide) in the gene. For example, if a gene is composed of 500 nucleotides, perhaps two different strains both have a mutation at nucleotide number 377 (from one end of the gene).

Application and Experimental Questions

E2. Electron microscopy must be used to visualize a virus.

E4. A P1 plaque mostly contains P1 bacteriophages that have a phage coat and P1 DNA. On occasion, however, a phage coat contains a segment of the bacterial chromosome. It would also contain material from the *E. coli* cells that had been lysed.

E6. In *E. coli* B, both *rII* mutants and wild-type strains could produce plaques. In *E. coli* K12(λ) only the rare recombinants that produced the wild-type gene would be able to form plaques. If the phage preparation used to infect *E. coli* B was not diluted enough, the entire bacterial lawn would be lysed, so it would be impossible to count the number of plaques.

E8. You would basically follow the strategy of Benzer (see Figure 10.11), except that you would not need to use different *E. coli* strains. Instead, you would grow the cells at different temperatures. You would coinfect an *E. coli* strain with two of the temperature-sensitive phages and then grow at 32°C. You would then reisolate phages, as shown in step 4 of Figure 10.11. You would take some of the phage preparation, dilute it greatly (10^{-8}), infect *E. coli*, and grow at 32°C. This would tell you how many total phages you had. You would also take some of the phage preparation, dilute it somewhat (10^{-6}), infect *E. coli*, and then grow at 37°C. At 37°C, only intragenic recombinants that produced wild-type phages would form plaques.

E10. You can first narrow down a mutation to one of several different regions by doing pairwise crosses with just a few deletion strains. After that, you make pairwise crosses only with strains that have also been narrowed down to the same region. You do not have to make pairwise crosses between all the mutant strains.

Questions for Student Discussion/Collaboration

2. Flower color in sweet peas is another example. When two different genes affect flower color, the white allele of one gene may be epistatic to the purple allele of another gene. A heterozygote for both genes would have purple flowers. In other words, the two purple alleles are complementing the two white alleles. Other similar examples are discussed in Chapter 5 in the section on Gene Interactions.

CHAPTER 11

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 11.1 In this experiment, the type R bacteria had taken up genetic material from the heat-killed type S bacteria, which converted the type R bacteria into type S. This enabled them to proliferate within the mouse and kill it.

FIGURE 11.2 RNase and protease were added to the DNA extract to rule out the possibility that small amounts of contaminating RNA or protein was responsible for converting the type R bacteria into type S.

FIGURE 11.4 The T2 genetic material must be injected into the cytoplasm of the bacterium for infection to occur.

FIGURE 11.5 In this experiment ^{32}P was used to label DNA and ^{35}S was used to label protein.

FIGURE 11.7 Ribose and uracil are not found in DNA.

FIGURE 11.10 Deoxyribose and phosphate form the backbone of a DNA strand.

FIGURE 11.11 Modeling is useful because it shows how molecules can fit together in a complicated three-dimensional structure.

FIGURE 11.14 Hydrogen bonding between base pairs and base stacking holds the DNA strands together.

FIGURE 11.15 The major and minor grooves are the indentations where the bases make contact with water. The major groove is wider than the minor groove.

FIGURE 11.16 B DNA is a right-handed helix and the backbone is helical, whereas Z DNA is a left-handed helix and the backbone appears to zigzag slightly. Z DNA has the bases tilted relative to the central axis, whereas they are perpendicular in B DNA. There are also minor differences in the number of bases pairs per turn.

FIGURE 11.17 Covalent bonds hold nucleotides together in an RNA strand.

FIGURE 11.18 A bonds with U and G bonds with C.

End-of-Chapter Questions:

Conceptual Questions

C2. The transformation process is described in Chapter 9.

1. A fragment of DNA binds to the cell surface.
2. It penetrates the cell wall/cell membrane.
3. It enters the cytoplasm.
4. It recombines with the chromosome.
5. The genes within the DNA are expressed (i.e., transcription and translation).
6. The gene products create a capsule. That is, they are enzymes that synthesize a capsule using cellular molecules as building blocks.

C4. The building blocks of a nucleotide are a sugar (ribose or deoxyribose), a nitrogenous base, and a phosphate group. In a nucleotide, the phosphate is already linked to the 5′ position on the sugar. When two nucleotides are hooked together, a phosphate on one nucleotide forms a covalent bond with the 3′ hyrdroxyl group on another nucleotide.

C6. The structure is a phosphate group connecting two sugars at the 3′ and 5′ positions, as shown in Figure 11.10.

C8. 3′–CCGTAATGTGATCCGGA–5′

C10. A drawing of a DNA helix with 10 bp per turn would look like Figure 11.14 in the textbook. To make 15 bp per turn, you would have to add 5 more base pairs, but the helix should still make only one complete turn.

C12. The bases occupy the major and minor grooves. Phosphate and sugar are found in the backbone. If a DNA-binding protein does not recognize a nucleotide sequence, it probably is not binding in the grooves, but instead is binding to the DNA backbone (i.e., sugar-phosphate sequence). DNA-binding proteins that recognize a base sequence must bind into a major or minor groove of the DNA, which is where the bases would be accessible to a DNA-binding protein. Most DNA-binding proteins that recognize a base sequence fit into the major groove. By comparison, other DNA-binding proteins such as histones, which do not recognize a base sequence, bind to the DNA backbone.

C14. The structure is shown in Figure 11.8. You begin numbering at the carbon that is to the right of the ring oxygen and continue to number the carbon atoms in a clockwise direction. Antiparallel means that the backbones are running in the opposite direction. In one strand, the sugar carbons are oriented in a 3′ to 5′ direction, but in the other strand they are oriented in a 5′ to 3′ direction.

C16. The structures are similar in that RNA and DNA double helices are helical, antiparallel, and base pairing is due to complementarity. The shapes of the helices are slightly different with regard to the number of base pairs per turn. Another difference is that RNA base pairing involves A to U, where DNA base pairing involves A to T.

C18. Its base sequence.

C20. G = 32%, C = 32%, A = 18%, T = 18%.

C22. Lysines and arginines, and also polar amino acids.

C24. They always run parallel.

C26. You would conclude that it is probably double-stranded RNA because the amount of A equals U and the amount of G equals C. Therefore, this molecule could be double-stranded and obey the AU/GC rule. However, it is also possible that it is merely a coincidence that A happens to equal U and G happens to equal C, and the genetic material is really single-stranded.

C28. There are 10^8 base pairs in this chromosome. In a double helix, a single base pair traverses about 0.34 nm, which equals 0.34×10^{-9} m. If we multiply the two values together:

$$10^8\,(0.34 \times 10^{-9}) = 0.34 \times 10^{-1}\text{ m, or } 0.034\text{ m, or } 3.4\text{ cm}$$

The answer is 3.4 cm, which equals 1.3 inches! That is enormously long considering that a typical human cell is only 10 to 100 μm in diameter. As described in Chapter 12, the DNA has to be greatly compacted to fit into a living cell.

C30. Yes, as long as some sequences are complementary and antiparallel to each other. It would be similar to the complementary double-stranded regions observed in RNA molecules (e.g., see Figures 11.18 and 11.19).

C32. Region 1 cannot form a stem-loop with region 2 and region 3 at the same time. Complementary regions of RNA form base pairs, not base triplets. The region 1/region 2 interaction would be slightly more stable than the region 1/region 3 interaction because it is one nucleotide longer, and it has a higher amount of GC base pairs. Remember that GC base pairs form three hydrogen bonds compared with AU base pairs, which form two hydrogen bonds. Therefore, helices with a higher GC content are more stable.

Application and Experimental Questions

E2. A. There are different possible reasons why most of the cells were not transformed.

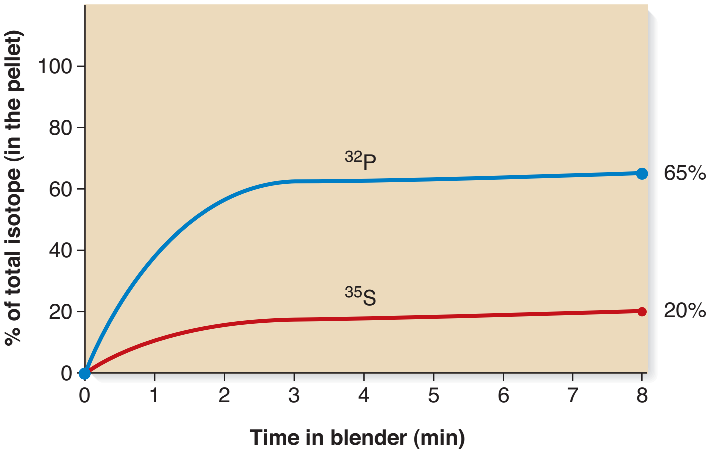
1. Most of the cells did not take up any of the type S DNA.
2. The type S DNA was usually degraded after it entered the type R bacteria.
3. The type S DNA was usually not expressed in the type R bacteria.

B. The antibody/centrifugation steps were used to remove the bacteria that had not been transformed. It enabled the researchers to determine the phenotype of the bacteria that had been transformed. If this step was omitted, there would have been so many colonies on the plate it would have been difficult to identify any transformed bacterial colonies, because they would have represented a very small proportion of the total number of bacterial colonies.

C. They were trying to demonstrate that it was really the DNA in their DNA extract that was the genetic material. It was possible that the extract was not entirely pure and could contain contaminating RNA or protein. However, treatment with RNase and protease did not prevent transformation, indicating that RNA and protein were not the genetic material. In contrast, treatment with DNase blocked transformation, confirming that DNA is the genetic material.

E4. A. There are several possible explanations why about 35% of the DNA is in the supernatant. One possibility is that not all of the DNA was injected into the bacterial cells. Alternatively, some of the cells may have been broken during the shearing procedure, thereby releasing the DNA.

B. If the radioactivity in the pellet had been counted instead of the supernatant, the following figure would be produced:



C. ^{32}P and ^{35}S were chosen as radioisotopes to label the phages because phosphorous is found in nucleic acids, and sulfur is found only in proteins.

D. There are multiple reasons why less than 100% of the phage protein is removed from the bacterial cells during the shearing process. For example, perhaps the shearing just is not strong enough to remove all of the phages, or perhaps the tail fibers remain embedded in the bacterium and only the head region is sheared off.

E6. This is really a matter of opinion. The Avery, MacLeod, and McCarty experiment seems to indicate directly that DNA is the genetic material, because DNase prevented transformation and RNase and protease did not. However, one could argue that the DNA is required for the rough bacteria to take up some other contaminant in the DNA preparation. It would seem that the other contaminant would not be RNA or protein. The Hershey and Chase experiments indicate that DNA is being injected into bacteria, although quantitatively the results are not entirely convincing. Some ^{35}S-labeled protein was not sheared off, so the results do not definitely rule out the possibility that protein could be the genetic material. But the results do indicate that DNA is the more likely candidate.

E8. A. The purpose of chromatography was to separate the different types of bases.

B. It was necessary to separate the bases and determine the total amount of each type of base. In a DNA strand, all of the bases are found within a single molecule, so it is difficult to measure the total amount of each type of base. When the bases are removed from the strand, each type can be purified, and then the total amount of each type of base can be measured by spectroscopy.

C. Chargaff's results would probably not be very convincing if done on a single species. The strength of his data was that all species appeared to conform to the AT/GC rule, suggesting that this is a consistent feature of DNA structure. In a single species, the observation that A = T and G = C could occur as a matter of chance.

Questions for Student Discussion/Collaboration

2. There are lots of possibilities. You could use a DNA-specific chemical and show that it causes heritable mutations. Perhaps you could inject an oocyte with a piece of DNA and produce a mouse with a new trait.

CHAPTER 12

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 12.1 The sequences of genes constitute most of a bacterial genome.

FIGURE 12.2 Three.

FIGURE 12.3 Proteins hold the bacterial chromosomal loops in place.

FIGURE 12.5 Strand separation is needed for certain processes such as DNA replication and RNA transcription.

FIGURE 12.6 ATP is needed so the DNA held in the upper jaws can pass through the break in the DNA and move to the region of the lower jaws.

FIGURE 12.7 Eukaryotes have centromeres and telomeres, which bacterial chromosomes do not. Also, eukaryotes typically have much more repetitive sequences.

FIGURE 12.8 One reason for variation in genome size is that the number of genes among different eukaryotes varies. A second reason is that the amount of repetitive sequences varies.

FIGURE 12.10 The nucleosome is 11 nm at its widest point.

FIGURE 12.13 The solenoid model depicts the nucleosomes in a repeating, spiral arrangement, whereas the zigzag model depicts a more irregular and dynamic arrangement of nucleosomes.

FIGURE 12.14 The nuclear matrix helps to organize and compact the chromosomes within the cell nucleus and also aids in their condensation during cell division.

FIGURE 12.15 A chromosome territory is a discrete region in the cell nucleus that is occupied by a single chromosome.

FIGURE 12.16 Active genes are found in more loosely packed euchromatic regions.

FIGURE 12.17 In the conversion of a 300-nm fiber to a 700-nm fiber, the radial loop domains become more tightly packed.

FIGURE 12.21 At the beginning of prophase, cohesin lines the region between sister chromatids. It is first removed from the arms of the sister chromatids but remains in the centromeric region. At anaphase, the cohesin at the centromeres is degraded, which allows the sister chromatids to separate.

End-of-Chapter Questions:

Conceptual Questions

C2. A bacterium with two nucleoids is similar to a diploid eukaryotic cell because it would have two copies of each gene. The bacterium is different, however, with regard to alleles. A eukaryotic cell can have two different alleles for the same gene. For example, a cell from a pea plant could be heterozygous, *Tt*, for the gene that affects height. By comparison, a bacterium with two nucleoids has two identical chromosomes. Therefore, a bacterium with two nucleoids is homozygous for its chromosomal genes. Note: As discussed in Chapter 9, a bacterium can contain another piece of DNA, called an F′ factor, that can carry a few genes. The alleles on an F′ factor can be different from the alleles on the bacterial chromosome.

C4. A. One loop is 40,000 bp. One base pair is 0.34 nm, which equals $0.34 \times 10^{-3}\ \mu m$. If we multiply the two together:

$$(40{,}000)(0.34 \times 10^{-3}) = 13.6\ \mu m$$

B. Circumference = πD

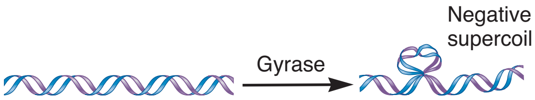
$13.6\ \mu m = \pi D$

$D = 4.3\ \mu m$

C. No, it is too big to fit inside of *E. coli*. Supercoiling is needed to make the loops more compact.

C6. These drugs would diminish the amount of negative supercoiling in DNA. Negative supercoiling is needed to compact the chromosomal DNA, and it also aids in strand separation. Bacteria might not be able to survive and/or transmit their chromosomes to daughter cells if their DNA was not compacted properly. Also, because negative supercoiling aids in strand separation, these drugs would make it more difficult for the DNA strands to separate. Therefore, the bacteria would have a difficult time transcribing their genes and replicating their DNA because both processes require strand separation. As discussed in Chapter 13, DNA replication is needed to make new copies of the genetic material to transmit from mother to daughter cells. If DNA replication was inhibited, the bacteria could not grow and divide into new daughter cells. As discussed in Chapters 14 through 16, gene transcription is necessary for bacterial cells to make proteins. If gene transcription was inhibited, the bacteria could not make many proteins that are necessary for survival.

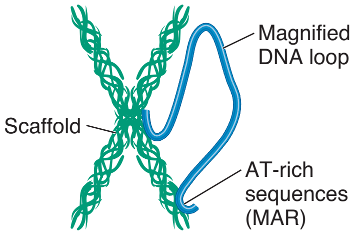
C8.



C10. The centromere is the attachment site for the kinetochore, which attaches to the spindle. If a chromosome is not attached to the spindle, it is free to “float around” within the cell, and it may not be near a pole when the nuclear membrane re-forms during telophase. If a chromosome is left outside of the nucleus, it is degraded during interphase. That is why the chromosome without a centromere may not be found in daughter cells.

C12. A nucleosome is composed of double-stranded DNA wrapped 1.65 times around an octamer of histones. In the 30-nm fiber, histone H1 helps to compact the nucleosomes. The solenoid and three-dimensional zigzag models describes how this compaction may occur in a 30-nm fiber. The solenoid model has a more regular, spiral arrangement of nucleosomes. By comparison, the arrangement of nucleosomes in the zigzag model appears somewhat random (zigzagging) and is thought to be more dynamic.

C14.



C16. During interphase, the chromosomes are found within the cell nucleus. They are less tightly packed and are transcriptionally active. Segments of chromosomes are anchored to the nuclear matrix. During M phase, the chromosomes become highly condensed, and the nuclear membrane is fragmented into vesicles. The chromosomal DNA remains anchored to a scaffold, formed from the nuclear matrix. The chromosomes eventually become attached to the spindle apparatus via microtubules that attach to the kinetochore, which is attached to the centromere.

C18. There are 146 bp around the core histones. If the linker region is 54 bp, we expect 200 bp of DNA (i.e., 146 + 54) for each nucleosome and linker region. If we divide 46,000 bp by 200 bp, we get 230. Because there are two molecules of H2A for each nucleosome, there would be 460 molecules of H2A in a 46,000-bp sample of DNA.

C20. The role of the core histones is to form the nucleosomes. In a nucleosome, the DNA is wrapped 1.65 times around the core histones. Histone H1 binds to the linker region. It may play a role in compacting the DNA into a 30-nm fiber.

C22. The answer is B and E. A Barr body is composed of a type of highly compacted chromatin called heterochromatin. Euchromatin is not so compacted. A Barr body is not composed of euchromatin. A Barr body is one chromosome, the X chromosome. The term *genome* refers to all the types of chromosomes that make up the genetic composition of an individual.

C24. SMC stands for *s*tructural *m*aintenance of *c*hromosomes. SMC proteins use energy from ATP to catalyze changes in chromosome structure. Together with topoisomerases, SMC proteins have been shown to promote major changes in DNA structure. Two examples of SMC proteins are condensin and cohesin, which play different roles in metaphase chromosome structure. The function of condensin is to promote the proper compaction of metaphase chromosomes, whereas that of cohesin is to promote the binding (i.e., cohesion) between sister chromatids.

Application and Experimental Questions

E2. Supercoiled DNA would look curled up into a relatively compact structure. You could add different purified topoisomerases and see how they affect the structure via microscopy. For example, DNA gyrase relaxes positive supercoils, whereas topoisomerase I relaxes negative supercoils. If we added topoisomerase I to a DNA preparation and it became less compacted, then the DNA was negatively supercoiled.

E4. With a moderate salt concentration, the nucleosome structure is still preserved so the same pattern of results would be observed. DNase I would cut the linker region and produce fragments of DNA that would be in multiples of 200 bp. However, with a high salt concentration, the core histones would be lost, and DNase I could cut anywhere. On the gel, you would see fragments of almost any size. Because there would be a continuum of fragments of many different sizes, the lane on the gel would probably look like a smear rather than having a few prominent bands of DNA.

E6. Lots of possibilities. You could digest it with DNase I and see if it gives multiples of 200 bp or so. You could try to purify proteins from the sample and see if eukaryotic proteins or bacterial proteins are present.

E8. A. Because the *Alu*I sequence is interspersed throughout all of the chromosomes, there would be many brightly colored spots along all chromosomes.

B. Only the centromeric region of the X chromosome would be brightly colored.

Questions for Student Discussion/Collaboration

2. This is a matter of opinion. It seems strange to have so much DNA that seems to have no obvious function. It's a waste of energy. Perhaps it has a function that we don't know about yet. On the other hand, evolution does allow bad things to accumulate within genomes such as genes that cause diseases, etc. Perhaps this is just another example of the negative consequences of evolution.

CHAPTER 13

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 13.1 The two features that allow DNA to be replicated are its double-stranded structure and the base pairing between A and T and between G and C.

FIGURE 13.2 The semiconservative model is correct.

FIGURE 13.5 The DnaA boxes are recognized by DnaA protein, which binds to them and causes the DNA to unwind at the AT-rich region.

FIGURE 13.6 Two replication forks are formed at the origin.

FIGURE 13.7 Primase is needed for DNA replication because DNA polymerase cannot initiate DNA replication on a bare template strand.

FIGURE 13.8 The template strand is read in the 3′ to 5′ direction.

FIGURE 13.10 The leading strand is made as one, long continuous strand in the same direction that the replication fork is moving. The lagging strand is made as Okazaki fragments in the direction away from the replication fork.

FIGURE 13.12 Having the replication machinery in a complex allows the coordination of different enzymes. For example, helicase and primase can work together to make multiple primers for the lagging strand.

FIGURE 13.14 Yes.

FIGURE 13.15 The oxygen in the newly made ester bond comes from the sugar.

FIGURE 13.18 Because eukaryotic chromosomes are so large, they need multiple origins so the DNA can be replicated in a reasonable length of time.

FIGURE 13.22 Six times (36 divided by 6).

FIGURE 13.23 An advantage of genetic recombination is that it may foster genetic diversity, which may produce organisms that have reproductive advantages over other organisms.

FIGURE 13.24 A heteroduplex region may be produced after branch migration because the two strands are not perfectly complementary.

FIGURE 13.25 During recombination, a D loop is formed when one DNA strand from one homolog invades the homologous region of the other homolog.

FIGURE 13.27 During gene conversion, the region of the gene that contained the genetic variation that created the *b* allele was digested away. The same region from the homolog, which carried the *B* allele, was then used as a template to make another copy of the *B* allele.

End-of-Chapter Questions:

Conceptual Questions

C2. Bidirectional replication refers to DNA replication in both directions starting from one origin.

C4. A. TTGGHTGUTGG
HHUUTHUGHUU

B. TTGGHTGUTGG
HHUUTHUGHUU

TTGGHTGUTGG CCAAACACCAA
AACCCACAACC HHUUTHUGHUU

TTGGHTGUTGG TTGGGTGTTGG CCAAACACCAA CCAAACACCAA
AACCCACAACC AACCCACAACC GGTTTGTGGTT HHUUTHUGHUU

C6. If we assume there are 4,600,000 bp of DNA, and that DNA replication is bidirectional at a rate of 750 nucleotides per second:

If there were just a single replication fork:

4,600,000/750 = 6,133 seconds, or 102.2 minutes

Because replication is bidirectional: 102.2/2 = 51.1 minutes

Actually, this is an average value based on a variety of growth conditions. Under optimal growth conditions, replication can occur substantially faster.

With regard to errors, if we assume an error rate of 1 mistake per 100,000,000 nucleotides:

4,600,000 × 1,000 bacteria = 4,600,000,000 nucleotides of replicated DNA

4,600,000,000/100,000,000 = 46 mistakes

When you think about it, this is pretty amazing. In this population, DNA polymerase would cause only 46 single mistakes in a total of 1000 bacteria, each containing 4.6 million bp of DNA.

C8. DNA polymerase would slide from right to left. The new strand would be

3′–CTAGGGCTAGGCGTATGTAAATGGTCTAGTGGTGG–5′

C10. A. When looking at Figure 13.5, the first, second, and fourth DnaA boxes are running in the same direction, and the third and fifth are running in the opposite direction. Once you realize that, you can see that the sequences are very similar to each other.

B. According to the direction of the first DnaA box, the consensus sequence is

TGTGGATAA
ACACCTATT

C. This sequence is nine nucleotides long. Because there are four kinds of nucleotides (i.e., A, T, G, and C), the chance of this sequence occurring by random chance is 4^{-9}, which equals once every 262,144 nucleotides. Because the *E. coli* chromosome is more than 10 times longer than this, it is fairly likely that this consensus sequence occurs elsewhere. There are not multiple origins, however, because the origin has five copies of the consensus sequence very close together. The chance of having five copies of this consensus sequence occurring close together (as a matter of random chance) is very small.

C12. 1. According to the AT/GC rule, a pyrimidine always hydrogen bonds with a purine. A transition still involves a pyrimidine hydrogen bonding to a purine, but a transversion causes a purine to hydrogen bond with a purine or a pyrimidine to hydrogen bond with a pyrimidine. The structure of the double helix makes it much more difficult for this latter type of hydrogen bonding to occur.

2. The induced-fit phenomenon of the active site of DNA polymerase makes it unlikely for DNA polymerase to catalyze covalent bond formation if the wrong nucleotide is bound to the template strand.

A transition mutation creates a somewhat bad interaction between the bases in opposite strands, but it is not as bad as the fit caused by a transversion mutation. In a transversion, a purine is opposite another purine, or a pyrimidine is opposite a pyrimidine. This is a very bad fit.

3. The proofreading function of DNA polymerase is able to detect and remove an incorrect nucleotide that has been incorporated into the growing strand. A transversion is going to cause a larger distortion in the structure of the double helix and make it more likely to be detected by the proofreading function of DNA polymerase.

C14. Primase and DNA polymerase are able to knock the single-strand binding proteins off the template DNA.

C16. A. The right Okazaki fragment was made first. It is farthest away from the replication fork. The fork (not seen in this diagram) would be to the left of the three Okazaki fragments, and moving from right to left.

B. The RNA primer in the right Okazaki fragment would be removed first. DNA polymerase would begin by elongating the DNA strand of the middle Okazaki fragment and remove the right RNA primer with its 5′ to 3′ exonuclease activity. DNA polymerase I would use the 3′ end of the DNA of the middle Okazaki fragment as a primer to synthesize DNA in the region where the right RNA primer is removed. If the middle fragment was not present, DNA polymerase could not fill in this DNA (because it needs a primer).

C. You need DNA ligase only at the right arrow. DNA polymerase I begins at the end of the left Okazaki fragment and synthesizes DNA to fill in the region as it removes the middle RNA primer. At the left arrow, DNA polymerase I is simply extending the length of the left Okazaki fragment. No ligase is needed here. When DNA polymerase I has extended the left Okazaki fragment through the entire region where the RNA primer has been removed, it hits the DNA of the middle Okazaki fragment. This occurs at the right arrow. At this point, the DNA of the middle Okazaki fragment has a 5′ end that is a monophosphate. DNA ligase is needed to connect this monophosphate with the 3′ end of the region where the middle RNA primer has been removed.

D. As mentioned in the answer to part C, the 5′ end of the DNA in the middle Okazaki fragment is a monophosphate. It is a monophosphate because it was previously connected to the RNA primer by a phosphoester bond. At the location of the right arrow, there was only one phosphate connecting this deoxyribonucleotide to the last ribonucleotide in the RNA primer. For DNA polymerase to function, the energy to connect two nucleotides comes from the hydrolysis of the incoming triphosphate. In this location shown at the right arrow, however, the nucleotide is already present at the 5′ end of the DNA, and it is a monophosphate. DNA ligase needs energy to connect this nucleotide with the left Okazaki fragment. It obtains energy from the hydrolysis of ATP or NAD^+.

C18. The picture would depict a ring of helicase proteins traveling along a DNA strand and separating the two helices, as shown in Figure 13.6.

C20. The leading strand is primed once, at the origin, and then DNA polymerase III synthesizes DNA continuously in the direction of the replication fork. In the lagging strand, many short pieces of DNA (Okazaki fragments) are made. This requires many RNA primers. The primers are removed by DNA polymerase I, which then fills in the gaps with DNA. DNA ligase then covalently connects the Okazaki fragments together. Having the enzymes within a complex such as a primosome or replisome provides coordination among the different steps in the replication process and thereby allows it to proceed faster and more efficiently.

C22. A processive enzyme is one that remains clamped to one of its substrates. In the case of DNA polymerase, it remains clamped to the template strand as it makes a new daughter strand. This is important to ensure a fast rate of DNA synthesis.

C24. The opposite strand is made in the conventional way by DNA polymerase using the strand made via telomerase as a template.

C26. The ends labeled B and C could not be replicated by DNA polymerase. DNA polymerase makes a strand in the 5′ to 3′ direction using a template strand that is running in the 3′ to 5′ direction. Also, DNA polymerase requires a primer. At the ends labeled B and C, there is no place (upstream) for a primer to be made.

C28. As shown in Figure 13.22, the first step involves a binding of telomerase to the telomere. The 3′ overhang binds to the complementary RNA in telomerase. For this reason, a 3′ overhang is necessary for telomerase to replicate the telomere.

C30. Branch migration will not create a heteroduplex during SCE because the sister chromatids are genetically identical. There should not be any mismatches between the complementary strands. Gene conversion cannot take place because the sister chromatids carry alleles that are already identical to each other.

C32. The two molecular mechanisms that can explain the phenomenon of gene conversion are DNA mismatch repair and gap repair synthesis. Both mechanisms could occur in the double-strand break model.

C34. A recombinant chromosome is one that has been derived from a crossover and contains a combination of alleles that is different from the parental chromosomes. A recombinant chromosome is a hybrid of the parental chromosomes.

C36. Gene conversion occurs when a pair of different alleles is converted to a pair of identical alleles. For example, a pair of *Bb* alleles could be converted to *BB* or *bb*.

C38. Gene conversion is likely to take place near the breakpoint. According to the double-strand break model, a gap may be created by the digestion of one DNA strand in the double helix. Gap repair synthesis may result in gene conversion. A second way that gene conversion can occur is by mismatch repair. A heteroduplex may be created after DNA strand migration and may be repaired in such a way as to cause gene conversion.

C40. No, it does not necessarily involve DNA mismatch repair; it could also involve gap repair synthesis. The double-strand break model involves a migration of DNA strands and the digestion of a gap. Therefore, in this gap region, only one chromatid is providing the DNA strands. As seen in Figure 13.25, the top chromosome is using the top DNA strand from the bottom chromosome in the gap region. The bottom chromosome is using the bottom DNA strand from the bottom chromosome. After DNA synthesis, both chromosomes may have the same allele.

Application and Experimental Questions

E2. A. You would probably still see a band of DNA, but you would only see a heavy band.

B. You would probably not see a band because the DNA would not be released from the bacteria. The bacteria would sediment to the bottom of the tube.

C. You would not see a band. UV light is needed to see the DNA, which absorbs light in the UV region.

E4. For a DNA strand to grow, a phosphoester bond is formed between the 3′ –OH group on one nucleotide and the innermost 5′ phosphate group on the incoming nucleotide (see Figure 13.15). If the –OH group is missing, a phosphoester bond cannot form.

Questions for Student Discussion/Collaboration

2. This is a matter of opinion. It needs to be fast so that life can occur on the time scale that we are accustomed to. The error-free issue is very important. Because DNA stores information, it is important for that information to be maintained with high fidelity. Otherwise, its storage ability is greatly compromised. Finally, the regulation of DNA replication is necessary to ensure the proper of amount of DNA per cell.

CHAPTER 14

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 14.2 The mutation would not affect the length of the RNA, because it would not terminate transcription. However, the encoded polypeptide would be shorter.

FIGURE 14.5 When comparing a group of related sequences, the consensus sequence is the most common base found at each location within that sequence.

FIGURE 14.6 Parts of sigma factor must fit into the major groove so it can recognize a base sequence of a promoter.

FIGURE 14.7 The –10 is AT-rich, which has fewer hydrogen bonds than a region with a lot of GC base pairs

FIGURE 14.10 The mutation would prevent ρ-dependent termination of transcription.

FIGURE 14.11 NusA helps RNA polymerase to pause, which facilitates transcriptional termination.

FIGURE 14.13 The TATA box provides a precise starting point for the transcription of eukaryotic structural genes.

FIGURE 14.14 The phosphorylation of CTD allows RNA polymerase to proceed to the elongation phase of transcription.

FIGURE 14.17 An endonuclease can cleave within a strand, whereas an exonuclease digests a strand, one nucleotide at a time, starting at one end.

FIGURE 14.18 Splicing via a spliceosome is very common in eukaryotes.

FIGURE 14.20 snRNPs are involved in recognizing the intron boundaries, cutting out the intron, and connecting the two adjacent exons together.

FIGURE 14.21 The 7-methylguanosine cap is important for the proper splicing of pre-mRNA, the exit of mRNA from the nucleus, and the binding of mRNA to a ribosome.

FIGURE 14.23 The consequence of RNA editing is that the amino acid sequence of an encoded polypeptide may be changed, which can affect its function.

End-of-Chapter Questions:

Conceptual Questions

C2. The release of sigma factor marks the transition to the elongation stage of transcription.

C4. GGCATTGTCA

C6. The most highly conserved positions are the first, second, and sixth. In general, when promoter sequences are conserved, they are more likely to be important for binding. That explains why changes are not found at these positions; if a mutation altered a conserved position, the promoter would probably not work very well. By comparison, changes are occasionally tolerated at the fourth position and frequently at the third and fifth positions. The positions that tolerate changes are less important for binding by sigma factor.

C8. This will not affect transcription. However, it will affect translation by preventing the initiation of polypeptide synthesis.

C10. Sigma factor can slide along the major groove of the DNA. In this way, it is able to recognize base sequences that are exposed in the groove. When it encounters a promoter sequence, hydrogen bonding between the bases and the sigma factor protein can promote a tight and specific interaction.

C12. DNA-G/RNA-C

DNA-C/RNA-G

DNA-A/RNA-U

DNA-T/RNA-A

The template strand is 3′–CCGTACGTAATGCCGTAGTGTGATCCCTAG–5′ and the coding strand is 5′–GGCATGCATTACGGCATCACACTAGGGATC–3′. The promoter would be to the left (in the 3′ direction) of the template strand.

C14. Transcriptional termination occurs when the hydrogen bonding is broken between the DNA and the part of the newly made RNA transcript that is located in the open complex.

C16. DNA helicase and ρ protein bind to a nucleic acid strand and travel in the 5′ to 3′ direction. When they encounter a double-stranded region, they break the hydrogen bonds between complementary strands. ρ protein is different from DNA helicase in that it moves along an RNA strand, whereas DNA helicase moves along a DNA strand. The purpose of DNA helicase function is to promote DNA replication, whereas the purpose of ρ protein function is to promote transcriptional termination.

C18. A. Mutations that alter the uracil-rich region by introducing G’s and C’s, and mutations that prevent the formation of the stem-loop structure.

B. Mutations that alter the termination sequence, and mutations that alter the ρ recognition site.

C. Eventually, somewhere downstream from the gene, another transcriptional termination sequence would be found, and transcription would terminate there. This second termination sequence might be found randomly or it might be at the end of an adjacent gene.

C20. Eukaryotic promoters are somewhat variable with regard to the pattern of sequence elements that may be found. In the case of structural genes that are transcribed by RNA polymerase II, it is common to have a TATA box, which is about 25 bp upstream from a transcriptional start site. The TATA box is important in the identification of the transcriptional start site and the assembly of RNA polymerase and various transcription factors. The transcriptional start site defines where transcription actually begins.

C22. The two models are described in Figure 14.15. The allosteric model is more like ρ-independent termination, whereas the torpedo model is more like ρ-dependent termination. In the torpedo model, a protein knocks RNA polymerase off the DNA, much like the effects of ρ protein.

C24. Hydrogen bonding is usually the predominant type of interaction when proteins and DNA follow an assembly and disassembly process. In addition, ionic bonding and hydrophobic interactions could occur. Covalent interactions would not occur. High temperature and high salt concentrations tend to break hydrogen bonds. Therefore, high temperature and high salt would inhibit assembly and stimulate disassembly.

C26. In bacteria, the 5′ end of the tRNA is cleaved by RNase P. The 3′ end is cleaved by a different endonuclease, and then a few nucleotides are digested away by an exonuclease that removes nucleotides until it reaches a CCA sequence.

C28. A ribozyme is an enzyme whose catalytic part is composed of RNA. Examples are RNase P and self-splicing group I and II introns. It is thought that the spliceosome may contain catalytic RNAs as well.

C30. Self-splicing means that an RNA molecule can splice itself without the aid of a protein. Group I and II introns can be self-splicing, although proteins can also enhance the rate of processing.

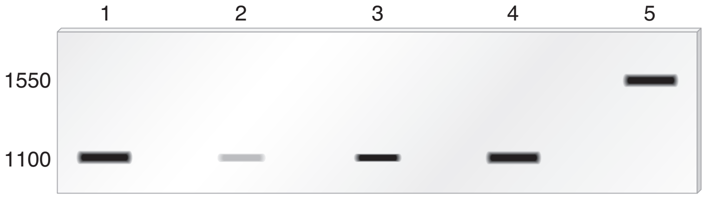
C32. In alternative splicing, variation occurs in the pattern of splicing so the resulting mRNAs contain alternative combinations of exons. The biological significance is that two or more different proteins can be produced from a single gene. This is a more efficient use of the genetic material. In multicellular organisms, alternative splicing is often used in a cell-specific manner.

C34. As shown at the left side of Figure 14.18, the guanosine, which binds to the guanosine-binding site, does not have a phosphate group attached to it. This guanosine is the nucleoside that winds up at the 5′ end of the intron. Therefore, the intron does not have a phosphate group at its 5′ end.

C36. U5.

Application and Experimental Questions

E2. The 1100-nucleotide band would be observed from a normal individual (lane 1). A deletion that removed the –50 to –100 region would greatly diminish transcription, so the homozygote would produce hardly any of the transcript (just a faint amount, as shown in lane 2), and the heterozygote would produce roughly half as much of the 1100-nucleotide transcript (lane 3) compared with a normal individual. A nonsense codon would not have an effect on transcription; it affects only translation. So the individual with this mutation would produce a normal amount of the 1100-nucleotide transcript (lane 4). A mutation that removed the splice acceptor site would prevent splicing. Therefore, this individual would produce a 1550-nucleotide transcript (actually, 1547 to be precise, 1550 minus 3). The Northern blot is shown here:



E4. A. It would not be retarded because ρ protein would not bind to the mRNA that is encoded by a gene that is terminated in a ρ-independent manner. The mRNA from such genes does not contain the sequence near the 3′ end that acts as a recognition site for the binding of ρ protein.

B. It would be retarded because ρ-protein would bind to the mRNA.

C. It would be retarded because U1 would bind to the pre-mRNA.

D. It would not be retarded because U1 would not bind to mRNA that has already had its introns removed. U1 binds only to pre-mRNA.

E6. A. mRNA molecules would bind to this column because they have a polyA tail. The string of adenine nucleotides in the polyA tail is complementary to stretch of thymine in the poly-dT column, so the two would hydrogen bond to each other. To purify mRNAs, one begins with a sample of cells; the cells need to be broken open by some technique such as homogenization or sonication. This would release the RNAs and other cellular macromolecules. The large cellular structures (organelles, membranes, etc.) could be removed from the cell extract by a centrifugation step. The large cellular structures would be found in the pellet, whereas soluble molecules such as RNA and proteins would stay in the supernatant. At this point, you would want the supernatant to contain a high salt concentration and neutral pH. The supernatant would then be poured over the poly-dT column. The mRNAs would bind to the poly-dT column, and other molecules (i.e., other types of RNAs and proteins) would flow through the column. Because the mRNAs would bind to the poly-dT column via hydrogen bonds, to break the bonds, you could add a solution that contains a low salt concentration and/or a high pH. This would release the mRNAs, which would then be collected in a low-salt/high-pH solution as it dripped from the column.

B. The basic strategy is to attach a short stretch of DNA nucleotides to the column matrix that is complementary to the type of RNA that you want to purify. For example, if an rRNA contained a sequence 5′–AUUCCUCCA–3′, a researcher could chemically synthesize an oligonucleotide with the sequence 3′–TAAGGAGGT–5′ and attach it to the column matrix. To purify rRNA, one would use this 3′–TAAGGAGGT–5′ column and follow the general strategy described in part A.

Questions for Student Discussion/Collaboration

2. RNA transcripts come in two basic types: those that function as RNA (e.g., tRNA, rRNA, etc.) versus those that are translated (i.e., mRNA). As described in this chapter, they play a myriad of functional roles. RNAs that form complexes with proteins carry out some interesting roles. In some cases, the role is to bind other types of RNA molecules. For example, rRNA in bacteria plays a role in binding mRNA. In other cases, the RNA plays a catalytic role. An example is RNaseP. The structure and function of RNA molecules may be enhanced by forming a complex with proteins.

CHAPTER 15

Note: The answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 15.1 Alkaptonuria occurs when homogentisic acid oxidase is defective.

FIGURE 15.2 The strain 2 mutants are unable to convert *O*-acetylhomoserine into cystathionine.

FIGURE 15.3 The role of DNA is to store the information that specifies the amino acid sequence of a polypeptide. It is a storage role.

FIGURE 15.5 Tryptophan and phenylalanine are the least soluble in water.

FIGURE 15.7 Hydrogen bonding promotes the formation of secondary structures in proteins.

FIGURE 15.9 Only one radiolabeled amino acid was found in each sample. If radioactivity was trapped on the filter, this meant that the codon triplet specified that particular amino acid.

FIGURE 15.10 A tRNA has an anticodon that recognizes a codon in the mRNA. It also has a 3′ acceptor site where the correct amino acid is attached.

FIGURE 15.12 A charged tRNA has an amino acid attached to it.

FIGURE 15.13 The wobble rules allow for a smaller population of tRNAs to recognize all of the possible mRNA codons.

FIGURE 15.15 A site in mRNA promotes the binding of the mRNA to the ribosome. The codons are needed during elongation to specify the polypeptide sequence. The stop codon is needed to terminate transcription.

FIGURE 15.17 The Shine-Dalgarno sequence in mRNA is complementary to a region in the 16S rRNA within the small ribosomal subunit. These complementary sequences hydrogen bond with each other.

FIGURE 15.18 Peptidyl transferase catalyzes the peptide bond formation between amino acids in the growing polypeptide chain.

FIGURE 15.19 The structures of release factors, which are proteins, resemble the structures of tRNAs.

End-of-Chapter Questions:

Conceptual Questions

C2. When we say the genetic code is degenerate, it means that more than one codon can specify the same amino acid. For example, GGG, GGC, GGA, and GGU all specify glycine. In general, the genetic code is nearly universal, because it is used in the same way by viruses, prokaryotes, fungi, plants, and animals. As discussed in Table 15.2, there are a few exceptions, which occur primarily in protists and yeast and mammalian mitochondria.

C4. A. This mutant tRNA would recognize glycine codons in the mRNA but would put in tryptophan amino acids where glycine amino acids are supposed to be in the polypeptide chain.

B. This mutation tells us that the aminoacyl-tRNA synthetase is primarily recognizing other regions of the tRNA molecule besides the anticodon region. In other words, tryptophanyl-tRNA synthetase (the aminoacyl-tRNA synthetase that attaches tryptophan) primarily recognizes other regions of the tRNA^{Trp} sequence (that is, other than the anticodon region). If aminoacyl-tRNA synthetases recognized only the anticodon region, we would expect glycyl-tRNA synthetase to recognize this mutant tRNA and attach glycine. That is not what happens.

C6. A. The answer is three. There are six leucine codons: UUA, UUG, CUU, CUC, CUA, and CUG. The anticodon AAU would recognize UUA and UUG. You would need two other tRNAs to efficiently recognize the other four leucine codons. These could be GAG and GAU or GAA and GAU.

B. The answer is one. There is only one codon, AUG, so you need only one tRNA with the anticodon UAC.

C. The answer is three. There are six serine codons: AGU, AGC, UCU, UCC, UCA, and UCG. You would need only one tRNA to recognize AGU and AGC. This tRNA could have the anticodon UCG or UCA. You would need two tRNAs to efficiently recognize the other four tRNAs. These could be AGG and AGU or AGA and AGU.

C8. 3′–CUU–5′ or 3′–CUC–5′

C10. It can recognize 5′–GGU–3′, 5′–GGC–3′, and 5′–GGA–3′. All of these specify glycine.

C12. All tRNA molecules have some basic features in common. They all have a cloverleaf structure with three stem-loop structures. The second stem-loop contains the anticodon sequence that recognizes the codon sequence in mRNA. At the 3′ end, there is an acceptor stem, with the sequence CCA, that serves as an attachment site for an amino acid. Most tRNAs also have base modifications that occur within their nucleotide sequences.

C14. The role of aminoacyl-tRNA synthetase is to specifically recognize tRNA molecules and attach the correct amino acid to them. This ability is sometimes described as the second genetic code because the specificity of the attachment is a critical step in deciphering the genetic code. For example, if a tRNA has a 3′–GGG–5′ anticodon, it will recognize a 5′–CCC–3′ codon, which should specify proline. It is essential that the aminoacyl-tRNA synthetase known as prolyl-tRNA-synthetase recognizes this tRNA and attaches proline to the 3′ end. The other aminoacyl-tRNA synthetases should not recognize this tRNA.

C16. Bases that have been chemically modified can occur at various locations throughout the tRNA molecule. The significance of all of these modifications is not entirely known. However, within the anticodon region, base modification alters base pairing to allow the anticodon to recognize two or more different bases within the codon.

C18. No, it is not. Due to the wobble rules, the 5′ base in the anticodon of a tRNA can recognize two or more bases in the third (3′) position of the mRNA. Therefore, any given cell type synthesizes far fewer than 61 types of tRNAs.

C20. The assembly process is very complex at the molecular level. In eukaryotes, 33 proteins and one rRNA assemble to form a 40S subunit, and 49 proteins and three rRNAs assemble to form a 60S subunit. This assembly occurs within the nucleolus.

C22. A. On the surface of the 30S subunit and at the interface between the two subunits

B. Within the 50S subunit C. From the 50S subunit

D. To the 30S subunit

C24. Most bacterial mRNAs contain a Shine-Dalgarno sequence, which is necessary for the binding of the mRNA to the small ribosomal subunit. This sequence, UUAGGAGGU, is complementary to a sequence in the 16S rRNA. Due to this complementarity, these sequences will hydrogen bond to each other during the initiation stage of translation.

C26. The ribosome binds at the 5′ end of the mRNA and then scans in the 3′ direction in search of an AUG start codon. If it finds one that reasonably obeys Kozak's rules, it will begin translation at that site. Aside from an AUG start codon, two other important features are a guanosine at the +4 position and a purine at the −3 position.

C28. The A (aminoacyl) site is the location where a tRNA carrying a single amino acid initially binds. The only exception is the initiator tRNA, which binds to the P (peptidyl) site. The growing polypeptide chain is removed from the tRNA in the P site and transferred to the amino acid attached to the tRNA in the A site. The ribosome translocates in the 3′ direction, with the result that the two tRNAs in the P and A sites are moved to the E (exit) and P sites, and the uncharged tRNA in the E site is released.

C30. The initiation phase involves the binding of the Shine-Dalgarno sequence to the rRNA in the 30S subunit. The elongation phase involves the binding of anticodons in tRNA to codons in mRNA.

C32. A. The E site and P sites. (Note: A tRNA without an amino acid attached is only briefly found in the P site, just before translocation occurs.)

B. P site and A site (Note: A tRNA with a polypeptide chain attached is only briefly found in the A site, just before translocation occurs.)

C. Usually the A site, except the initiator tRNA, which can be found in the P site.

C34. The tRNAs bind to the mRNA because their anticodon and codon sequences are complementary. When the ribosome translocates in the 5′ to 3′ direction, the tRNAs remain bound to their complementary codons, and the two tRNAs shift from the A site and P site to the P site and E site. If the ribosome moved in the 3′ direction, it would have to dislodge the tRNAs and drag them to a new position where they would not (necessarily) be complementary to the mRNA.

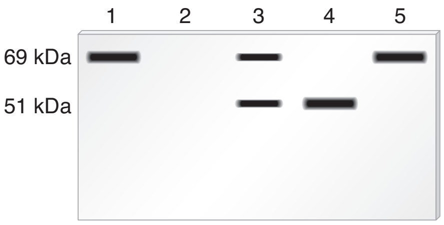
C36. 52.

Application and Experimental Questions

E2. The initiation phase of translation is very different in bacteria and in eukaryotes, so they would not be translated very efficiently. A bacterial mRNA would not be translated very efficiently in a eukaryotic translation system, because it lacks a cap structure attached to its 5′ end. A eukaryotic mRNA would not have a Shine-Dalgarno sequence near its 5′ end, so it would not be translated very efficiently in a bacterial translation system.

E4. Looking at the figure, the 5′ end of the template DNA strand is toward the right side. The 5′ ends of the mRNAs are farthest from the DNA, and the 3′ ends of the mRNAs are closest to the DNA. The start codons are slightly downstream from the 5′ ends of the RNAs.

E6.



E8. A. If codon usage were significantly different between kangaroo and yeast cells, the translation process would be inhibited. For example, if the preferred leucine codon in kangaroos was CUU, translation would probably be slow in a yeast translation system. We would expect the cell-free translation system from yeast cells to primarily contain leucine tRNAs with an anticodon sequence that is AAC, because this tRNA^{Leu} would match the preferred yeast leucine codon, which is UUG. In a yeast translation system, there probably would not be a large amount of tRNA with an anticodon of GAA, which would match the preferred leucine codon, CUU, of kangaroos. For this reason, kangaroo mRNA would not be translated very well in a yeast translation system, but it probably would be translated to some degree.

B. The advantage of codon bias is that a cell can rely on a smaller population of tRNA molecules to efficiently translate its proteins. A disadvantage is that mutations, which do not change the amino acid sequence but do change a codon (e.g., UUG to UUA), may inhibit the production of a polypeptide if a preferred codon is changed to a nonpreferred codon.

Questions for Student Discussion/Collaboration

2. This could be a very long list. There are similarities along several lines:

   1. There is a lot of molecular recognition going on, either between two nucleic acid molecules or between proteins and nucleic acid molecules. Students may see these as similarities or differences, depending on their point of view.
   2. There is biosynthesis going on in both processes. Small building blocks are being connected together. This requires an input of energy.
   3. Genetic signals determine the beginning and ending of these processes.

   There are also many differences:

   1. Transcription produces an RNA molecule with a similar structure to the DNA, whereas translation produces a polypeptide with a structure that is very different from RNA.
   2. Depending on your point of view, it seems that translation is more biochemically complex, requiring more proteins and RNA molecules to accomplish the task.

CHAPTER 16

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 16.1 Gene regulation is more efficient. A cell does not was energy making RNAs and proteins it does not need.

FIGURE 16.2 Activators and repressors are regulatory proteins. Inducers, corepressors, and inhibitors are small effector molecules.

FIGURE 16.3 The *lacZ*, *lacY*, and *lacA* genes are under the control of the *lac* promoter.

FIGURE 16.5 The lac repressor is bound to the *lac* operon when it is not exposed to lactose—when allolactose is not bound to the repressor.

FIGURE 16.8 The repressor protein allows the cell to avoid turning on the operon in the absence of lactose. The activator protein allows the cell to choose between glucose and lactose.

FIGURE 16.9 The data in which O_2 and O_3 are deleted indicate that O_1, by itself, is not very strong in repressing the *lac* operon.

FIGURE 16.11 Tryptophan acts as a corepressor that causes the trp repressor to bind to the *trp* operon and repress transcription.

FIGURE 16.12 Hydrogen bonding between complementary sequences causes the stem-loops to form.

FIGURE 16.13 The presence of tryptophan causes the *trpL* gene to be translated to its stop codon. This blocks both regions 1 and 2, which allows a 3-4 stem-loop to form. The 3-4 stem-loop causes transcriptional termination.

FIGURE 16.14 The *micF* antisense RNA binds to the *ompF* mRNA and inhibits its translation.

FIGURE 16.15 Feedback inhibition prevents a bacterium from over producing the product of a metabolic pathway.

FIGURE 16.16 The RNA conformation with an antiterminator stem-loop favors transcription.

FIGURE 16.17 The RNA conformation with the Shine-Dalgarno antisequestor favors translation.

End-of-Chapter Questions:

Conceptual Questions

C2. In bacteria, gene regulation greatly enhances the efficiency of cell growth. It takes a lot of energy to transcribe and translate genes. Therefore, a cell is much more efficient and better at competing in its environment if it expresses genes only when the gene product is needed. For example, a bacterium will express only the genes that are necessary for lactose metabolism when a bacterium is exposed to lactose. When the environment is missing lactose, these genes are turned off. Similarly, when tryptophan levels are high within the cytoplasm, the genes that are required for tryptophan biosynthesis are repressed.

C4. A. Regulatory protein
B. Effector molecule
C. DNA segment
D. Effector molecule
E. Regulatory protein
F. DNA segment
G. Effector molecule

C6. A mutation that has a *cis*-effect is within a genetic regulatory sequence, such as an operator site, that affects the binding of a genetic regulatory protein. A *cis*-effect mutation affects only the adjacent genes that the genetic regulatory sequence controls. A mutation having a *trans*-effect is usually in a gene that encodes a genetic regulatory protein. A *trans*-effect mutation can be complemented in a merozygote experiment by the introduction of a normal gene that encodes the regulatory protein.

C8. A. No transcription would take place. The *lac* operon could not be expressed.

B. No regulation would take place. The operon would be continuously turned on.

C. The rest of the operon would function normally but none of the transacetylase would be made.

C10. Diauxic growth refers to the phenomenon in which a cell first uses up one type of sugar (e.g., glucose) before it begins to metabolize a second sugar (e.g., lactose). In this case, it is caused by gene regulation. When a bacterial cell is exposed to both sugars, the uptake of glucose causes the cAMP levels in the cell to fall. When this occurs, the catabolite activator protein (CAP) is removed from the *lac* operon so it is not able to be activated by CAP.

C12. A mutation that prevented the lac repressor from binding to the operator would make the *lac* operon constitutive only in the absence of glucose. However, this mutation would not be entirely constitutive because transcription would be inhibited in the presence of glucose. The disadvantage of constitutive expression of the *lac* operon is that the bacterial cell would waste a lot of energy transcribing the genes and translating the mRNA when lactose was not present.

C14. A. Attenuation will not occur because loop 2–3 will form.

B. Attenuation will occur because 2–3 cannot form, so 3–4 will form.

C. Attenuation will not occur because 3–4 cannot form.

D. Attenuation will not occur because 3–4 cannot form.

C16. The addition of Gs and Cs into the U-rich sequence would prevent attenuation. The U-rich sequence promotes the dissociation of the mRNA from the DNA, when the terminator stem-loop forms. This causes RNA polymerase to dissociate from the DNA, thereby causing transcriptional termination. The UGGUUGUC sequence would probably not dissociate because of the Gs and Cs. Remember that GC base pairs have three hydrogen bonds and are more stable than AU base pairs, which have only two hydrogen bonds.

C18. It takes a lot of cellular energy to translate mRNA into a protein. A cell wastes less energy if it prevents the initiation of translation rather than a later stage such as elongation or termination.

C20. One mechanism is that histidine could act as corepressor that shuts down the transcription of the histidine synthetase gene. A second mechanism would be that histidine could act as an inhibitor via feedback inhibition. A third possibility is that histidine inhibits the ability of the mRNA encoding histidine synthetase to be translated. Perhaps it induces a gene that encodes an antisense RNA. If the amount of histidine synthetase protein was identical in the presence and absence of extracellular histidine, a feedback inhibition mechanism is favored, because this affects only the activity of the histidine synthetase enzyme, not the amount of the enzyme. The other two mechanisms would diminish the amount of this protein.

C22. The two proteins are similar in that both bind to a segment of DNA and repress transcription. They are different in three ways. (1) They recognize different effector molecules (i.e., the lac repressor recognizes allolactose, and the trp repressor recognizes tryptophan. (2) Allolactose causes the lac repressor to release from the operator, whereas tryptophan causes the trp repressor to bind to its operator. (3) The sequences of the operator sites that these two proteins recognize are different from each other. Otherwise, the lac repressor could bind to the *trp* operator, and the trp repressor could bind to the *lac* operator.

Application and Experimental Questions

E2. In samples loaded in lanes 1 and 4, we expect the repressor to bind to the operator because no lactose is present. In the sample loaded into lane 4, the CAP protein could still bind cAMP because there is no glucose. However, there really is no difference between lanes 1 and 4, so it does not look like the CAP can activate transcription when the lac repressor is bound. If we compare samples loaded into lanes 2 and 3, the lac repressor would not be bound in either case, and the CAP would not be bound in the sample loaded into lane 3. There is less transcription in lane 3 compared to lane 2, but because there is some transcription seen in lane 3, we can conclude that the removal of the CAP (because cAMP levels are low) is not entirely effective at preventing transcription. Overall, the results indicate that the binding of the lac repressor is more effective at preventing transcription of the *lac* operon compared to the removal of the CAP.

E4. A. Yes, if you do not sonicate, then β-galactosidase will not be released from the cell, and not much yellow color will be observed. (Note: You may observe a little yellow color because some of β-ONPG may be taken into the cell.)

B. No, you should still get yellow color in the first two tubes even if you forgot to add lactose because the unmated strain does not have a functional lac repressor.

C. Yes, if you forgot to add β-ONPG, you could not get yellow color because the cleavage of β-ONPG by β-galactosidase is what produces the yellow color.

E6. You could mate a strain that has an F′ factor carrying a normal *lac* operon and a normal *lacI* gene to this mutant strain. Because the mutation is in the operator site, you would still continue to get expression of β-galactosidase, even in the absence of lactose.

Questions for Student Discussion/Collaboration

2. A DNA loop may inhibit transcription by preventing RNA polymerase from recognizing the promoter. Or, it may inhibit transcription by preventing the formation of the open complex. Alternatively, a bend may enhance transcription by exposing the base sequence that the sigma factor of RNA polymerase recognizes. The bend may expose the major groove in such a way that this base sequence is more accessible to binding by sigma factor and RNA polymerase.

CHAPTER 17

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 17.1 Transcriptional regulation is the most energy-efficient, because a cell avoids wasting energy making RNA or protein.

FIGURE 17.3 An α helix can bind into the major groove of DNA and recognize a specific sequence of bases.

FIGURE 17.4 If TFIID cannot bind to the TATA box, RNA polymerase will not be recruited to the core promoter, and therefore transcription will not begin.

FIGURE 17.5 When an activator interacts with mediator, it causes mediator to phosphorylate CTD, which causes RNA polymerase to proceed to the elongation phase of transcription.

FIGURE 17.7 The glucocorticoid receptor binds only next to genes that have a GRE by their core promoters.

FIGURE 17.8 Nucleosome eviction may allow certain proteins access to binding to particular sites in the DNA.

FIGURE 17.9 Histone modifications may directly affect the interaction between histones and the DNA, or they may affect the binding of other proteins to the chromatin.

FIGURE 17.10 An NFR is needed at the core promoter so that transcription factors can recognize enhancers and so the preinitiation complex can form.

FIGURE 17.11 Histone eviction or displacement is needed for elongation because RNA polymerase cannot transcribe though nucleosomes. It needs to unwind the DNA for transcription to take place.

FIGURE 17.13 In part (a), DNA methylation is preventing an activator protein from binding to the DNA. This prevents transcriptional activation.

FIGURE 17.14 *De novo* methylation occurs on unmethylated DNA, whereas maintenance methylation occurs on hemimethylated DNA.

FIGURE 17.15 Insulators prevent one gene from regulating a neighboring gene. This allows each gene to control its own regulation.

FIGURE 17.18 Exon 3 would be spliced out. Exon 4 would still be within the mRNA.

FIGURE 17.20 If multiple copies of a gene are inserted into a genome, it becomes more likely that an endogenous promoter will transcribe the antisense strand of that gene.

FIGURE 17.22 A RISC binds to a specific mRNA because the small RNA within RISC is complementary to that mRNA. The bonding is hydrogen bonding between complementary bases.

FIGURE 17.24 The mutation would cause the over production of ferritin, because ferritin synthesis would occur even if iron levels were low.

End-of-Chapter Questions:

Conceptual Questions

C2. Regulatory elements are relatively short genetic sequences that are recognized by regulatory transcription factors. After the regulatory transcription factor has bound to the regulatory element, it will affect the rate of transcription, either activating it or repressing it, depending on the action of the regulatory protein. Regulatory elements are typically located in the upstream region near the promoter, but they can be located almost anywhere (i.e., upstream and downstream) and even quite far from the promoter.

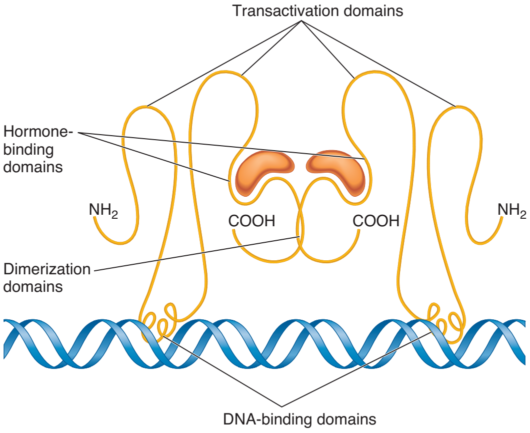
C4. Transcriptional activation occurs when a regulatory transcription factor binds to a regulatory element and activates transcription. Such proteins, called activators, may interact with TFIID and/or mediator to promote the assembly of RNA polymerase and general transcription factors at the promoter region. They also could alter the structure of chromatin so that RNA polymerase and transcription factors are able to gain access to the promoter. Transcriptional inhibition occurs when a regulatory transcription factor inhibits transcription. Such repressors also may interact with TFIID and/or mediator to inhibit RNA polymerase.

C6. A. DNA binding. B. DNA binding. C. Protein dimerization.

C8.
1. It could be in the DNA-binding domain, so that the receptor would not recognize GREs.
2. It could be in the HSP90 domain, so that HSP90 would not be released when the hormone binds.
3. It could be in the dimerization domain, so that the receptor would not dimerize.
4. It could be in the nuclear localization domain, so that the receptor would not travel into the nucleus.
5. It could be in the domain that activates RNA polymerase, so that the receptor would not activate transcription, even though it could bind to GREs.

C10. Possibility 2 is the correct one. Because we already know that the E protein and the Id protein form heterodimers with myogenic bHLH, we expect all three proteins to have a leucine zipper. Leucine zippers promote dimer formation. We also need to explain why the Id protein inhibits transcription, but the E protein enhances transcription. As seen in possibility 2, the Id protein does not have a DNA-binding domain. Therefore, if it forms a heterodimer with myogenic bHLH, the heterodimer probably will not bind to the DNA very well. In contrast, when the E protein forms a heterodimer with myogenic bHLH, there will be two DNA-binding domains, which would promote good binding to the DNA.

C12. This is one hypothetical drawing of the glucocorticoid receptor. As discussed in your textbook, the glucocorticoid receptor forms a homodimer. The dimer shown here has rotational symmetry. If you flip the left side around, it has the same shape as the right side. The hormone-binding, DNA-binding, dimerization, and transactivation domains are labeled. The glucocorticoid hormone is shown in orange.



C14. A histone variant is a histone with an amino acid sequence that is slightly different from a core histone. Histone variants play specialized roles with regard to chromatin structure and function.

C16. The histone code hypothesis is the idea that the patterns of covalent modifications of histones are recognized by proteins that affect chromatin structure and function. In this way, the modification of histones plays a role in gene regulation.

C18. A potentially harmful consequence would be that transcription may be initiated at multiple points within a gene, thereby producing many nonfunctional transcripts. This would be waste of energy.

C20. DNA methylation is the attachment of a methyl group to a base within the DNA. In many eukaryotic species, this occurs on cytosine at a CG sequence. After *de novo* methylation has occurred, it is passed from mother to daughter cell. Because DNA replication is semiconservative, the newly made DNA contains one strand that is methylated and one that is not. DNA methyltransferase recognizes this hemimethylated DNA and methylates the cytosine in the unmethylated DNA strand; this event is called maintenance methylation.

C22. A CpG island is a stretch of 1000 to 2000 bp in length that contains a high number of CpG sites. CpG islands are often located near promoters. When the island is methylated, transcription is inhibited. This inhibition may be the result of the inability of the transcriptional activators to recognize the methylated promoter and/or the effects of methyl-CpG-binding proteins, which may promote a closed chromatin conformation.

C24. As shown in Figure 17.17, the unique feature of the smooth muscle mRNA for α-tropomyosin is that it contains exon 2. Splicing factors that are only found in smooth muscle cells may recognize the splice junction at the 3′ end of intron 1 and the 5′ end of intron 2 and promote splicing at these sites. This would cause exon 2 to be included in the mRNA. Furthermore, because smooth muscle mRNA does not contain exon 3, a splicing suppressor may bind to the 3′ end of intron 2. This would promote exon skipping, so that exon 3 would not be contained in the mRNA.

C26. A cell may need to respond rather quickly to a toxic substance in order to avoid cell death. Translational and posttranslational mechanisms are much faster than transcriptional activation, in which it is necessary to up-regulate the gene, synthesize the mRNA, and then translate the mRNA to make a functional protein.

C28. RNA interference refers to the phenomenon in which the presence of a double-stranded RNA molecule leads to the silencing of a complementary mRNA. For this to occur, the double-stranded RNA is processed by Dicer into small RNA fragments (miRNA or siRNA). The miRNA or siRNA associates with a complex called the RNA-induced silencing complex (RISC). The RISC complex then binds to a complementary mRNA. This either leads to mRNA degradation or inhibits translation.

C30. If mRNA stability is low, it is degraded more rapidly. Therefore, low stability results in a low mRNA concentration. The length of the polyA tail is one factor that affects stability. A longer tail makes mRNA more stable. Certain mRNAs have sequences that affect their half-lives. For example, AU-rich elements (AREs) are found in many short-lived mRNAs. The AREs are recognized by cellular proteins that cause the mRNAs to be rapidly degraded.

Application and Experimental Questions

E2. If the DNA band is 3800 bp long, it means that the site is unmethylated, because it has been cut by *Not*I (and *Not*I cannot cut methylated DNA). If the DNA fragment is 5300 bp long, the DNA is not cut by *Not*I, so we assume that it is methylated. Now let's begin our interpretation of these data with lane 4. When the gene is isolated from root tissue, the DNA is not methylated because it runs at 3800 bp. This suggests that gene *T* is expressed in root tissue. In the other samples, the DNA runs at 5300 bp, indicating that the DNA is methylated. The pattern of methylation seen here is consistent with the known function of gene *T*. We would expect it to be expressed in root cells, because it functions in the uptake of phosphate from the soil. It would be silenced in the other parts of the plant via methylation.

E4. A. Based on the transformation of kidney cells, a silencer is in region B. Based on the transformation of pancreatic cells, an enhancer is in region A.

B. The pancreatic cells must not express the repressor protein that binds to the silencer in region B.

C. The kidney cells must express a repressor protein that binds to region B and represses transcription. Kidney cells express the downstream gene only if the silencer is removed. As mentioned in part B, this repressor is not expressed in pancreatic cells.

E6. When they injected just antisense *mex-3* RNA, they observed lower but detectable levels of *mex-3* mRNA. However, the injection of both sense and antisense RNA, which would form a double-stranded structure, resulted in a complete loss of *mex-3* mRNA in the cells. In this way, double-stranded RNA would silence the expression of the *mex-3* gene.

Questions for Student Discussion/Collaboration

2. Probably the most efficient method would be to systematically make deletions of progressively smaller sizes. For example, you could begin by deleting 20,000 bp on either side of the gene and see if that affects transcription. If you found that only the deletion on the 5′ end of the gene had an effect, you could then start making deletions from the 5′ end, perhaps in 10,000-bp or 5000-bp increments until you localized regulatory elements. You would then make smaller deletions in the putative region until it was down to a hundred or a few dozen nucleotides. At this point, you might conduct site-directed mutagenesis, as described in Chapter 19, as a way to specifically identify the regulatory element sequence.

CHAPTER 18

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 18.2 Position effect means that the expression of a gene has been altered due to a change in its position along a chromosome.

FIGURE 18.3 In part (b), the sequence of the eye color gene is not changed. The change in expression is due to a position effect.

FIGURE 18.5 A somatic mutation is not passed from parent to offspring.

FIGURE 18.6 Random mutation hypothesis.

FIGURE 18.7 The colonies on the secondary plates appearing in the same locations indicates that the mutations occurred on the master plate before exposure to T1. Therefore, the mutations did not occur as a physiological adaptation to T1 exposure.

FIGURE 18.8 The probability is 75%.

FIGURE 18.9 The deamination of methylcytosine is more difficult to repair because thymine is a normal base found in DNA. This makes it difficult for DNA repair enzymes to distinguish between the correct and altered strand.

FIGURE 18.11 A reactive oxygen species is a chemical that contains oxygen and may react with cellular molecules in a way that may cause harm.

FIGURE 18.14 5-Bromouracil causes a transition.

FIGURE 18.15 Thymine dimer formation is often the result of exposure to UV light. It most commonly occurs in skin cells.

FIGURE 18.16 Enzymes within the rat liver extract may convert nonmutagenic molecules into a mutagenic form. Adding the extract allows researchers to identify molecules that may be mutagenic in people.

FIGURE 18.17 Photolyase is particularly valuable to plants, which are exposed to sun and susceptible to thymine dimer formation.

FIGURE 18.19 Cuts are made on both sides of the DNA damage so the region can be removed by UvrD.

FIGURE 18.20 MutH distinguishes between the parental strand and the newly made daughter strand, which ensures that the daughter strand is repaired.

FIGURE 18.21 MutH ensures that the daughter strand is removed.

FIGURE 18.23 An advantage of this DNA repair system is that it can occur at any stage of the cell cycle. A disadvantage is that it may be imprecise and result in a short deletion in the DNA.

FIGURE 18.24 One advantage of specifically targeting a DNA repair system to actively transcribing genes is that it promotes the repair of DNA regions that are functionally important. Also, actively transcribed genes may be more susceptible to mutation.

End-of-Chapter Questions:

Conceptual Questions

C2. It is a gene mutation, a point mutation, a base substitution, a transition mutation, a deleterious mutation, a mutant allele, a nonsense mutation, a conditional mutation, and a temperature-sensitive lethal mutation.

C4. A. It would probably inhibit protein function, particularly if it was not near the end of the coding sequence.

B. It may or may not affect protein function, depending on the nature of the amino acid substitution and whether the substitution is in a critical region of the protein.

C. It would increase the amount of functional protein.

D. It may affect protein function if the alteration in splicing changes an exon in the mRNA that results in a protein with a perturbed structure.

C6. A. Not appropriate, because the second mutation is at a different codon.

B. Appropriate.

C. Not appropriate, because the second mutation is in the same gene as the first mutation.

D. Appropriate.

C8. A. Silent, because the same amino acid (glycine) is encoded by GGA and GGT.

B. Missense, because a different amino acid is encoded by CGA compared to GGA.

C. Missense, because a different amino acid is encoded by GTT compared to GAT.

D. Frameshift, because an extra base is inserted into the sequence.

C10. One possibility is that a translocation may move a gene next to a heterochromatic region of another chromosome, thereby diminishing its expression, or it could be moved next to a euchromatic region and increase its expression. Another possibility is that the translocation breakpoint may move the gene next to a new promoter or regulatory sequences that may now influence the gene's expression.

C12. A. No, the position (i.e., chromosomal location) of a gene has not been altered.

B. Yes, the expression of a gene has been altered because it has been moved to a new chromosomal location.

C. Yes, the expression of a gene has been altered because it has been moved to a new chromosomal location.

C14. If a mutation within the germ line is passed to an offspring, all the cells of the offspring's body will carry the mutation. A somatic mutation affects only the somatic cell in which it originated and all of the daughter cells that the somatic cell produces. If a somatic mutation occurs early during embryonic development, it may affect a fairly large region of the organism. Because germ-line mutations affect the entire organism, they are potentially more harmful (or beneficial), but this is not always the case. Somatic mutations can cause quite harmful effects such as cancer.

C16. A thymine dimer can interfere with DNA replication because DNA polymerase cannot slide past the dimer and add bases to the newly growing strand. Alkylating mutagens such as nitrous acid will cause DNA replication to make mistakes in the base pairing. For example, an alkylated cytosine will base-pair with adenine during DNA replication, thereby creating a mutation in the newly made strand. A third example is 5-bromouracil, which is a thymine analog. It may base-pair with guanine instead of adenine during DNA replication.

C18. During TNRE, a trinucleotide repeat sequence gets longer. If someone was mildly affected with a TNRE disorder, he or she might be concerned that an expansion of the repeat might occur during gamete formation, yielding offspring more severely affected with the disorder, a phenomenon called anticipation. This phenomenon may depend on the sex of the parent with the TNRE.

C20. According to random mutation theory, spontaneous mutations can occur in any gene and do not involve exposure of the organism to a particular environment that selects for specific types of mutation. However, the structure of chromatin may cause certain regions of the DNA to be more susceptible to random mutations. For example, DNA in an open conformation may be more accessible to mutagens and more likely to incur mutations. Simlarly, hot spots—certain regions of a gene that are more likely to mutate than other regions—can occur within a single gene. Also, another reason that some genes mutate at a higher rate is that some genes are larger than others, which provides a greater chance for mutation.

C22. Excision repair systems could fix this damage. Also, homologous recombination repair could fix the damage.

C24. Anticipation means that the TNRE expands even further in future generations. Anticipation may depend on the sex of the parent with the TNRE.

C26. The mutation frequency is the total number of mutant alleles divided by the total number of alleles in the population. If there are 1,422,000 babies, there are 2,844,000 copies of this gene (because each baby has 2 copies). The mutation frequency is 31/2,844,000, which equals 1.09×10^{-5}. The mutation rate is the number of new mutations per generation. There are 13 babies who did not have a parent with achondroplasia; thus, 13 is the number of new mutations. If we calculate the mutation rate as the number of new mutations in a given gene per generation, then we should divide 13 by 2,844,000. In this case, the mutation rate would be 4.6×10^{-6}.

C28. The effects of mutations are cumulative. If one mutation occurs in a cell, this mutation will be passed to the daughter cells. If a mutation occurs in the daughter cell, now there will be two mutations. These two mutations will be passed to the next generation of daughter cells, and so forth. The accumulation of many mutations eventually kills the cells. That is why mutagens are more effective at killing dividing cells than nondividing cells. It is because the number of mutations accumulates to a lethal level.

There are two main side effects to this treatment. First, some normal (noncancerous) cells of the body, particularly skin cells and intestinal cells, are actively dividing. These cells are also killed by chemotherapy and radiation therapy. Secondly, it is possible that the therapy may produce mutations that will cause noncancerous cells to become cancerous. For these reasons, there is a maximal dose of chemotherapy or radiation therapy that is recommended.

C30. A. Yes.

B. No, the albino trait affects the entire individual.

C. No, the early apple-producing trait affects the entire tree.

D. Yes.

C32. Mismatch repair is aimed at eliminating mismatches that may have occurred during DNA replication. In this case, the wrong base is in the newly made strand. The binding of MutH, which occurs on a hemimethylated sequence, provides a sensing mechanism to distinguish between the unmethylated and methylated strands. In other words, MutH binds to the hemimethylated DNA in a way that allows the mismatch repair system to distinguish which strand is methylated and which is not.

C34. Because sister chromatids are genetically identical, an advantage of homologous recombination is that it can be an error-free mechanism to repair a DSB. A disadvantage, however, is that it occurs only during the S and G_2 phase of the cell cycle in eukaryotes or following DNA replication in bacteria. An advantage of NHEJ is that it doesn't involve the participation of a sister chromatid, so it can occur at any stage of the cell cycle. However, a disadvantage is that NHEJ can result in small deletions in the region that has been repaired. Overall, NHEJ is a quick but error-prone repair mechanism, whereas HR is a more accurate method of repair that is limited to certain stages of the cell cycle.

C36. In *E. coli*, the TRCF recognizes when RNA polymerase is stalled on the DNA. This stalling may be due to DNA damage such as a thymine dimer. The TRCF removes RNA polymerase and recruits the excision DNA repair system to the region, thereby promoting the repair of the template strand of DNA. It is beneficial for preferentially repairing actively transcribed DNA because it is functionally important. It is a DNA region that encodes a gene.

C38. The underlying genetic defect that causes xeroderma pigmentosum is a defect in one of the genes that encode a polypeptide involved with nucleotide excision repair. These individuals are defective in repairing DNA abnormalities such as thymine dimers and abnormal bases. Therefore, they are very sensitive to environmental agents such as UV light. Because they are defective at repair, UV light is more likely to cause mutations in these people than in unaffected individuals. For this reason, people with XP develop pigmentation abnormalities and premalignant lesions, and have a high predisposition to skin cancer.

C40. Both types of repair systems recognize an abnormality in the DNA and excise the abnormal strand. The normal strand is then used as a template to synthesize a complementary strand of DNA. The systems differ in the types of abnormalities they detect. The mismatch repair system detects base pair mismatches, whereas the excision repair system recognizes thymine dimers, chemically modified bases, missing bases, and certain types of cross-links. The mismatch repair system operates immediately after DNA replication, allowing it to distinguish between the daughter strand (which contains the wrong base) and the parental strand. The excision repair system can operate at any time in the cell cycle.

Application and Experimental Questions

E2. When cells from a master plate were replica-plated onto two plates containing selective media with the T1 phage, T1-resistant colonies were observed at the same locations on both plates. These results indicate that the mutations occurred randomly while on the master plate (in the absence of T1) rather than occurring as a result of exposure to T1. In other words, mutations are random events, and selective conditions may promote the survival of mutant strains that occur randomly.

To show that antibiotic resistance is due to random mutation, one could follow the same basic strategy except the secondary plates would contain the antibiotic instead of T1 phage. If the antibiotic resistance arose as a result of random mutation on the master plate, one would expect the antibiotic-resistant colonies to appear at the same locations on two different secondary plates.

E4. You would expose the bacteria to the physical agent. You could also expose the bacteria to the rat liver extract, but it is probably not necessary for two reasons. First, a physical mutagen is not something that a person would eat. Therefore, the actions of digestion via the liver are probably irrelevant if you are concerned that the agent might be a mutagen. Second, the rat liver extract would not be expected to alter the properties of a physical mutagen.

E6. The results suggest that the strain is defective in excision repair. If we compare the normal and mutant strains that have been incubated for 2 hours at 37°C, much of the radioactivity in the normal strain has been transferred to the soluble fraction because it has been excised. In the mutant strain, however, less of the radioactivity has been transferred to the soluble fraction, suggesting that it is not as efficient at removing thymine dimers.

Questions for Student Discussion/Collaboration

2. The worst time to be exposed to mutagens is at very early stages of embryonic development. An early embryo is most sensitive to mutation because it will affect a large region of the body. Adult must also worry about mutagens for several reasons. Mutations in somatic cells can cause cancer, a topic that is discussed in Chapter 23. Also, adults should be careful to avoid mutagens that may affect the ovaries or testes because these mutations could be passed to offspring.

CHAPTER 19

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 19.1 Eight.

FIGURE 19.2 In this experiment, the selectable marker gene selects for the growth of bacteria that have taken up a plasmid.

FIGURE 19.3 Reverse transcriptase refers to the idea that this enzyme catalyzes the opposite of transcription. It uses an RNA template to make DNA, whereas during transcription a DNA template is used to make RNA.

FIGURE 19.5 The products that predominate are the DNA fragments that are flanked by primers. This occurs because these products can be used as templates to make products only like themselves.

FIGURE 19.7 The probe must be cleaved to separate the quencher from the reporter.

FIGURE 19.9 The advantage of a DNA library is that it lacks introns. This is useful if a researcher wants to focus her or his attention on the coding

sequence of a gene or if she or he wants to express the gene in cells that do not splice out introns.

FIGURE 19.11 Conditions of low stringency reveal gene families.

FIGURE 19.13 The secondary antibody is labeled, which provides a way to detect the protein of interest.

FIGURE 19.15 The binding of a protein to DNA prevents DNase I from cleaving the region where the protein is bound.

FIGURE 19.18 Some examples are (1) to study the core promoter of a gene; (2) to study regulatory elements; (3) to identify splice sites; and (4) to study the importance of particular amino acids with regard to protein function.

End-of-Chapter Questions:

Conceptual Questions

C2. A restriction enzyme recognizes and binds to a specific DNA sequence and then cleaves a (covalent) phosphoester bond in each of two DNA strands.

C4. The term *cDNA* refers to DNA that is made using RNA as the starting material. Compared with genomic DNA, it lacks introns.

Application and Experimental Questions

E2. Remember that AT base-pairs form two hydrogen bonds, but GC base-pairs form three hydrogen bonds. The order (from stickiest to least sticky) is:

*Bam*HI = *Pst*I = *Sac*I > *Eco*RI > *Cla*I

E4. In conventional gene cloning, many copies are made because the vector replicates to a high copy number within the cell, and the cells divide to produce many more cells. In PCR, the replication of the DNA to produce many copies is facilitated by primers, deoxyribonucleoside triphosphates (dNTPs), and *Taq* polymerase.

E6. A recombinant vector is a vector that has a piece of "foreign" DNA inserted into it. The foreign DNA came from somewhere else, such as the chromosomal DNA of some organism. To construct a recombinant vector, the vector and source of foreign DNA are digested with the same restriction enzyme. The complementary ends of the fragments are allowed to hydrogen bond to each other (i.e., sticky ends are allowed to bind), and then DNA ligase is added to create covalent phosphoester bonds. In some cases, a piece of the foreign DNA becomes ligated to the vector, thereby creating a recombinant vector. In other cases, the two ends of the vector ligate back together, restoring the vector to its original structure.

As described in Figure 19.2, the insertion of foreign DNA can be detected using X-Gal. As seen here, the insertion of the foreign DNA causes the inactivation of the *lacZ* gene. The *lacZ* gene encodes the enzyme β-galactosidase, which converts the colorless compound X-Gal to a blue compound. If the *lacZ* gene is inactivated by the insertion of foreign DNA, the enzyme will not be produced, and the bacterial colonies will be white. If the vector has simply recircularized, and the *lacZ* gene remains intact, the enzyme will be produced, and the colonies will be blue.

E8. A thermostable form of DNA polymerase (e.g., *Taq* polymerase) is used in PCR because each PCR cycle involves a heating step to denature the DNA. This heating step would inactivate most forms of DNA polymerase. However, *Taq* polymerase is thermostable and can remain functional after many cycles of heating and cooling. It is not necessary to use a thermostable form of DNA polymerase in the techniques of dideoxy sequencing or site-directed mutagenesis. In these methods, DNA polymerase can be added after the annealing step, and the sample can be incubated at a temperature that does not inactivate most forms of DNA polymerase.

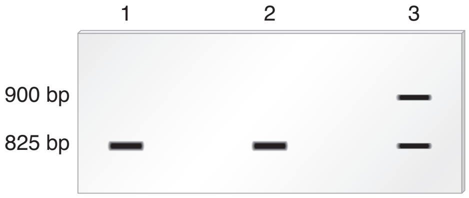
E10. The detector is a molecule whose fluorescence increases as the PCR product increases. One example is TaqMan. It has a fluorescent molecule and a quencher, which are connected by a primer that binds to the DNA that is amplified. When PCR is occurring, *Taq* polymerase digests the primer, thereby separating the fluorescent molecule from the quencher. Therefore, the fluorescence increases as the PCR product increases. Alternatively, other detectors intercalate into the DNA, which causes their fluorescence to increase. In this case, as the PCR product increases, the fluorescence also increases because the detectors are intercalating into the newly made DNA.

E12. One interpretation would be that the gene is part of a gene family. In this case, the family would contain four homologous members. At high stringency, the probe binds only to the gene that is its closest match, but at low stringency it recognizes the three other homologous genes.

E14. It would be necessary to use cDNA so the gene would not carry any introns. Bacterial cells do not contain spliceosomes (which are described in Chapter 14). To express a eukaryotic protein in bacteria, a researcher would clone cDNA into bacteria, because the cDNA does not contain introns.

E16. The purpose of gel electrophoresis is to separate the many DNA fragments, RNA molecules, or proteins that were obtained from the sample you want to analyze. This separation is based on molecular mass and allows you to identify the molecular mass of the DNA fragment, RNA molecule, or protein that is being recognized by the probe.

E18. The Northern blot is shown here. The female mouse expresses the same total amount of this mRNA as the male. In the heterozygous female, there would be 50% of the 900-bp band and 50% of the 825-bp band.



E20. Restriction enzymes recognize many sequences throughout the chromosomal DNA. If two fragments from different samples have the same molecular mass in a Southern blot, it is likely (though not certain) that the two fragments are found at the same chromosomal site in the genome. In this Southern blot, most of the transposable elements are found at the same sites within the genomes of these different yeast strains. However, a couple of bands are different among the three strains. These results indicate that the *Ty* element may occasionally transpose to a new location or that chromosomal changes (point mutations, chromosomal rearrangements, deletions, etc.) may have slightly changed the genomes among these three strains of yeast in a way that changes the distances between the restriction sites.

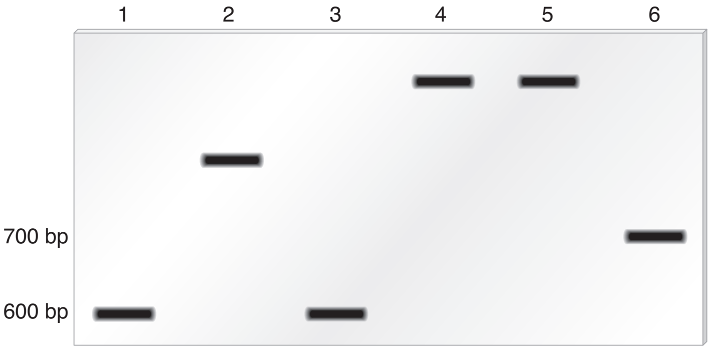
E22. Lane 1 shows that β-globin is made in normal red blood cells. In red blood cells from a thalassemia patient (lane 2), however, very little is made. Perhaps this person is homozygous for a down promoter mutation, which diminishes the transcription of the gene. As shown in lanes 3 and 4, β-globin is not made in muscle cells.

E24. Western blotting.

E26. You would first make a DNA library using chromosomal DNA from pig cells. You would then radiolabel a portion of a strand from the human β-globin gene and use it as a probe in a colony hybridization experiment; each colony would contain a different cloned piece of the pig genome. You would identify "hot" colonies that hybridize to the human β-globin probe. You would then go back to the master plate and pick these colonies and grow them in a test tube. You would then isolate the plasmid DNA from the colonies and subject the DNA to DNA sequencing. By comparing the DNA sequences of the human β-globin gene and the putative clones, you could determine if the putative clones were homologous to the human clone and likely to be the pig homolog of the β-globin gene.

E28. The rationale behind a gel retardation assay is that a segment of DNA with a protein bound to it will migrate more slowly through a gel than will the same DNA without any bound protein. A shift in a DNA band to a higher molecular mass provides a way to identify DNA-binding proteins.

E30. The glucocorticoid receptor will bind to a GRE if glucocorticoid hormone is also present (lane 2). The glucocorticoid receptor does not bind without hormone (lane 1) and it does not bind to a CRE (lane 6). The CREB protein will bind to a CRE with or without hormone (lanes 4 and 5), but it will not bind to a GRE (lane 3). The expected results are shown here. In this drawing, the binding of CREB protein to the 700-bp fragment results in a complex with a higher mass than the glucocorticoid receptor binding to the 600-bp fragment.



E32. The rationale behind a footprinting experiment has to do with accessibility. If a protein is bound to the DNA, it will cover up the part of the DNA where it is bound. This region of the DNA will be inaccessible to the actions of chemicals or enzymes that cleave the DNA, such as DNase I.

E34. AGGTCGGTTGCCATCGCAATAATTTCTGCCTGAACCCAATA

E36. There are lots of different strategies one could follow. For example, you could mutate every other base and see what happens. It would be best to make very nonconservative mutations such as a purine for a pyrimidine or a pyrimidine for a purine. If the mutation prevents protein binding in a gel retardation assay, then the mutation is probably within the regulatory element. If the mutation has no effect on protein binding, it probably is outside the regulatory element.

Questions for Student Discussion/Collaboration

2. A. Does a particular amino acid within a protein sequence play a critical role in the protein's structure or function?
   B. Does a DNA sequence function as a promoter?
   C. Does a DNA sequence function as a regulatory site?
   D. Does a DNA sequence function as a splicing junction?
   E. Is a sequence important for correct translation?
   F. Is a sequence important for RNA stability?
   AND MANY OTHERS. . .

CHAPTER 20

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 20.1 Cyanogen bromide separates β-galactosidase from the A or B chains.

FIGURE 20.3 In gene addition, a cloned gene is added to an organism's genome. In gene replacement, a cloned gene is swapped with a gene that is already present in the organism's genome.

FIGURE 20.5 The *Neo*^{R} gene selects for cells that have taken up the cloned gene.

FIGURE 20.6 A gene knockout replaces a normal gene with a gene that is inactivated. A gene knockin inserts a gene into a noncritical site in the genome.

FIGURE 20.7 The β-lactoglobulin promoter is used because it is expressed in mammary cells.

FIGURE 20.8 The nucleus of the oocyte is removed so the resulting organism contains (nuclear) genetic material only from the somatic cell.

FIGURE 20.9 No. Carbon Copy has not received genetic material from a different species.

FIGURE 20.10 When stem cells divide, they produce one cell that remains a stem cell and another cell that differentiates. This pattern maintains a population of stem cells.

FIGURE 20.12 Multipotent.

FIGURE 20.14 Only the T-DNA within the T-DNA vector is transferred to a plant.

FIGURE 20.17 The liposome method is relatively safe in that it doesn't evoke an immune response. A disadvantage is that it may be inefficient at getting the cloned gene into many cells. The retrovirus method is typically very good at getting the cloned gene into cells. A disadvantage is that it may cause a potentially harmful immune response.

End-of-Chapter Questions:

Conceptual Questions

C2. *A. radiobacter* synthesizes an antibiotic that kills *A. tumefaciens*. The genes, which are necessary for antibiotic biosynthesis and resistance, are plasmid-encoded and can be transferred during interspecies conjugation. If *A. tumefaciens* received this plasmid during conjugation, it would be resistant to killing. Therefore, the conjugation-deficient strain prevents the occurrence of *A. tumefaciens*–resistant strains.

C4. A biological control agent is an organism that prevents the harmful effects of some other agent in the environment. Examples include *Bacillus thuringiensis*, a bacterium that synthesizes proteins that act as toxins to kill insects and the use of *Agrobacterium radiobacter* to prevent crown gall disease caused by *A. tumefaciens*.

C6. A mouse model is a strain of mice that carries a mutation in a mouse gene that is analogous to a mutation in a human gene that causes disease. These mice can be used to study the disease and to test potential therapeutic agents.

C8. The T DNA gets transferred to the plant cell; it then is incorporated into the plant cell's genome.

C10. A. With regard to maternal-effect genes, the phenotype would depend on the animal that donated the egg. It is the cytoplasm of the egg that accumulates the gene products of maternal-effect genes.
B. The extranuclear traits depend on the mitochondrial genome. Mitochondria are found in the egg and in the somatic cell. So, theoretically, both cells could contribute extranuclear traits. In reality, however, researchers have found that the mitochondria in Dolly were from the animal that donated the egg. It is not clear why she had no mitochondria from the mammary cell.
C. The cloned animal would be genetically identical to the animal that donated the nucleus with regard to traits that are determined by nuclear genes, which are expressed during the lifetime of the organism. The cloned animal would/could differ from the animal that donated the nucleus with regard to traits that are determined by maternal-effect genes and extracytoplasmic genes. Such an animal is not a true clone, but it is likely that it would greatly resemble the animal that donated the nucleus, because the vast majority of genes are found in the cell nucleus.

C12. Some people are concerned with the release of genetically engineered microorganisms into the environment. The fear is that such organisms may continue to proliferate and it may not be possible to "stop them." A second concern involves the use of genetically engineered organisms in the food we eat. Some people are worried that genetically engineered organisms may pose an unknown health risk. A third issue is ethics. Some people feel that it is morally wrong to tamper with the genetics of organisms. This opinion may also apply to genetic techniques such as cloning, stem cell research, and gene therapy.

Application and Experimental Questions

E2. One possibility is to clone the toxin-producing genes from *B. thuringiensis* and introduce them into *P. syringae*. This bacterial strain would have the advantage of not needing repeated applications. However, it would be a recombinant strain and might be viewed in a negative light by people who are hesitant to use recombinant organisms in the field. By comparison, *B. thuringiensis* is a naturally occurring species.

E4. Basically, one can follow the strategy described in Figure 20.5. If homologous recombination occurs, only the *Neo*^{R} gene is incorporated into the genome. The cells will be neomycin-resistant and also resistant to gancyclovir. If gene addition occurs, the cells will usually be sensitive to gancyclovir. By growing the cells in the presence of neomycin and gancyclovir, one can select for homologous recombinants. The chimeras are identified because they have a mixture of light and dark fur.

E6.
1. *A. tumefaciens* (T DNA-mediated gene transfer): Genes are cloned into the T DNA of the Ti plasmid. This T DNA from this plasmid is transferred to the plant cell when it is infected by the bacterium.
2. Biolistic gene transfer: DNA is coated on microprojectiles that are "shot" into the plant cell.
3. Microinjection: A microscopic needle containing a solution of DNA is used to inject DNA into plant cells.
4. Electroporation: An electrical current is used to introduce DNA into plant cells.
5. Also, DNA can be introduced into protoplasts by treatment with polyethylene glycol and calcium phosphate. This method is similar to bacterial transformation procedures.

E8. In Mendel's work, and the work of many classical geneticists, an altered (mutant) phenotype is the initial way to identify a gene. For example, Mendel recognized a gene that affects plant height by the identification of tall and dwarf plants. The transmission of this gene could be followed in genetic crosses, and eventually, the gene could be cloned using molecular techniques. Reverse genetics uses the opposite sequence of steps. The gene is cloned first, and a phenotype for the gene is discovered later, by making a transgenic animal with a gene knockout.

E10. A chimera is an organism that is composed of cells from two different individuals (usually of the same species). Chimeras are made by mixing together embryonic cells from two individuals and allowing the cells to organize themselves and develop into a single individual.

E12. You would use Southern blotting to determine the number of gene copies. As described in Chapter 19, you will observe multiple bands in a Southern blot if there are multiple copies of a gene. You need to know this information to predict the outcome of crosses. For example, if there are four integrated copies at different sites in the genome, an offspring could inherit anywhere from zero to four copies of the gene. You would want to understand this so you could predict the phenotypes of the offspring.

You would use Northern blotting or Western blotting to monitor the level of gene expression. A Northern blot will indicate if the gene is transcribed into mRNA, and a Western blot will indicate if the mRNA is translated into protein.

E14. Reproductive cloning means the cloning of entire multicellular organisms. In plants, this is easy. Most species of plants can be cloned by asexual cuttings. In animals, cloning occurs naturally, as in identical twins. Identical twins are genetic replicas of each other because they begin from the same fertilized egg. (Note: There could be some somatic mutations that occur in identical twins that would make them slightly different.) Recently, as in the case of Dolly, reproductive cloning has become possible by fusing somatic cells with enucleated eggs. The advantage, from an agricultural point of view, is that reproductive cloning could allow one to choose the best animal in a herd and make many clones from it. Breeding would no longer be necessary. Also, breeding may be less reliable because the offspring inherit traits from both the mother and father.

E16. Ex vivo therapy involves the removal of living cells from the body and their modification after they have been removed. The modified cells are then reintroduced back into a person's body. This approach works well for cells such as blood cells that are easily removed and replaced. By comparison, this approach would not work very well for many other cell types. For example, lung cells cannot be removed and put back again. In this case, in vivo approaches must be sought.

Questions for Student Discussion/Collaboration

2. This is a matter of opinon. Examples include general traits such as size, productivity, flavor, disease resistance, pesticide resistance, and many more. Several techniques, which are discussed in this chapter, could be used to produce transgenic plants and animals.

CHAPTER 21

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 21.1 A genetic map is a chart that describes the order and relative distances between genes and other sites along a chromosome.

FIGURE 21.2 The probe binds only to a specific site that has a complementary sequence.

FIGURE 21.4 An RFLP is a polymorphism in the lengths of DNA fragments that are generated when chromosomal DNA is digested with a restriction enzyme.

FIGURE 21.5 Microsatellites are polymorphic when the number of repeat sequences varies.

FIGURE 21.7 A contig is a series of clones that have overlapping fragments of DNA that span a region of chromosome or an entire chromosome.

FIGURE 21.10 Their advantage is they can be used to clone large fragments of DNA.

FIGURE 21.12 Chromosome walking is used to identify a gene that has already been mapped to a site along a chromosome.

FIGURE 21.14 Whole-genome shotgun sequencing is now more common because DNA sequencing has become relatively inexpensive.

FIGURE 21.15 Yes. The sequence is determined as the DNA is synthesized.

FIGURE 21.16 Replicative transposition and retrotransposition always cause the TE to increase in number. Simple transposition, by itself, does not increase the TE number. However, as discussed later, simple transposition can increase the numbers of a TE if it occurs around the time of DNA replication.

FIGURE 21.20 Reverse transcriptase uses RNA as a template to make a strand of DNA.

End-of-Chapter Questions:

Conceptual Questions

C2. A. Yes. B. No, this is only one chromosome in the genome.
C. Yes. D. Yes.

C4. The ends of a short region would be flanked by direct repeats. This is a common characteristic of transposable elements. In addition, many elements contain IRs or LTRs that are involved in the transposition process. One might also look for the presence of a transposase or reverse transcriptase gene, although this is not an absolute requirement, because nonautonomous transposable elements typically lack transposase or reverse transcriptase.

C6. Retroelements have the greatest potential for proliferation because the element is transcribed into RNA as an intermediate. Many copies of this RNA could be transcribed and then copied into DNA by reverse transcriptase. Theoretically, many copies of the element could be inserted into the genome in a single generation.

C8. Keep in mind that each type of transposase recognizes only the inverted repeats of a particular type of transposable element. The mosquitoes must express a transposase that recognizes the Z elements. This explains why the Z elements are mobile. This Z element transposase must not recognize the inverted repeats of the X elements; the inverted repeats of the X elements and Z elements must have different sequences. This same group of mosquitoes must not express a transposase that recognizes the X elements. This explains why the X elements are very stable.

C10. A transposon is excised as a segment of DNA, and then it transposes to a new location via transposase. A retroelement is transcribed as RNA, and then reverse transcriptase makes a double-stranded copy of DNA. Integrase then inserts this DNA copy into the chromosome. All transposable elements have direct repeats. Transposons that move as DNA have inverted repeats, and they may encode transposase (as well as other genes). Retroelements do not have inverted repeats, but they may or may not have LTRs. Autonomous retroelements encode reverse transcriptase and integrase.

C12. A deletion may result when recombination occurs between two different elements that align in the same direction. The region between the elements is deleted. An inversion can happen when recombination occurs between two different elements that are oriented in opposite directions. A translocation can result when recombination occurs between transposable elements that are located on different (nonhomologous) chromosomes. In other words, the TEs on different chromosomes align themselves and a crossover occurs.

Application and Experimental Questions

E2. They are complementary to each other.

E4. Because normal cells contain two copies of chromosome 14, one would expect that a probe would bind to complementary DNA sequences on both of these chromosomes. If a probe recognized only one of two chromosomes, this means that one of the copies of chromosome 14 has been lost, or it has suffered a deletion in the region where the probe binds. With regard to cancer, the loss of this genetic material may be related to the uncontrollable cell growth.

E6. After the cells and chromosomes have been fixed to the slide, it is possible to add two or more different probes that recognize different sequences (i.e., different sites) within the genome. Each probe has a different fluorescence emission wavelength. Usually, a researcher will use computer imagery that recognizes the wavelength of each probe and then assigns that probe a bright color. The color seen by the researcher is not the actual color emitted by the probe; it is a secondary color assigned by the computer. In a sense, the probes, with the aid of a computer, are "painting" the regions of the chromosomes that are recognized by a probe. An example of chromosome painting is shown in Figure 21.3. In this example, human chromosome 5 is painted with six different colors.

E8. A contig is a collection of clones that contain overlapping segments of DNA that span a particular region of a chromosome. To determine if two clones are overlapping, one could conduct a Southern blotting experiment. In this approach, one of the clones is used as a probe. If it is overlapping with the second clone, it will bind to it in a Southern blot. Therefore, the second clone is run on a gel and the first clone is used as a probe. If the band corresponding to the second clone is labeled, this means that the two clones are overlapping.

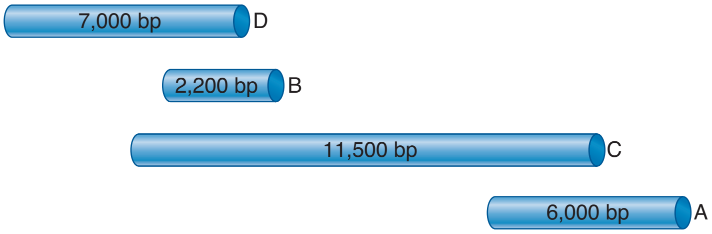
E10. BAC cloning vectors have the replication properties of a bacterial chromosome and the cloning properties of a plasmid. To replicate like a chromosome, the BAC vector contains an origin of replication from an F factor. Therefore, in a bacterial cell, a BAC can behave as a chromosome. Like a plasmid, BACs also contain selectable markers and convenient cloning sites for the insertion of large segments of DNA. The primary advantage is the ability to clone very large pieces of DNA.

E12. Child 1 and child 3 belong to father 2; child 2 and child 4 belong to father 1; child 5 could belong to either father.

E14. An explanation is that the rate of recombination between homologous chromosomes is different during oogenesis compared to spermatogenesis. Physical mapping measures the actual distance (in base pairs) between markers. The physical mapping of chromosomes in males and females reveals that they are the same lengths. Therefore, the sizes of chromosomes in males and females are the same. The differences obtained in linkage maps are due to differences in the rates of recombination during oogenesis versus spermatogenesis.

E16. Besides a selectable marker and an origin that will replicate in *E. coli*, YAC vectors also require two telomere sequences: a centromere sequence, and an ARS sequence. The telomeres are needed to prevent the shortening of the artificial chromosome from the ends. The centromere sequence is needed for the proper segregation of the artificial chromosome during meiosis and mitosis. The ARS sequence is the yeast equivalent of an origin of replication, which is needed so the YAC DNA can be replicated.

E18. Note that the insert of cosmid B is contained completely within the insert of cosmid C.



E20. A. The general strategy is shown in Figure 21.12. The researcher begins at a certain location and then walks toward the gene of interest. You begin with a clone that has a marker that is known to map relatively close to the gene of interest. A piece of DNA at the end of the insert is subcloned and then used in a Southern blot to identify an adjacent clone in a cosmid DNA library. This is the first "step." The end of this clone is subcloned to make the next step. And so on. Eventually, after many steps, you will arrive at your gene of interest.

B. In this example, you would begin at STS-3. If you walked a few steps and happened upon STS-2, you would know that you were walking in the wrong direction.

C. This is a difficult aspect of chromosome walking. Basically, you would walk toward gene X using DNA from a normal individual and DNA from an individual with a mutant gene X. When you have found a site where the sequences are different between the normal and mutant individual, you may have found gene X. You would eventually have to confirm this by analyzing the DNA sequence of this region and determining that it encodes a functional gene.

E22. In the hierarchical approach, the genome is mapped to create a YAC or BAC contig for each chromosome. The inserts within each contig are then subjected to shotgun sequencing. By comparison, in the whole-genome shotgun-sequencing approach, the mapping step is eliminated. Instead, the whole genome is chopped up into small pieces, which are sequenced from both ends. This latter method is quicker because it eliminates the mapping step, which can be quite time-consuming.

E24. Sequence by synthesis is a form of next-generation DNA sequencing in which the base sequence is determined as the DNA strand is being made. Pyrosequencing is an example.

Questions for Student Discussion/Collaboration

2. This is a matter of opinion. Many people would say that the ability to identify many human genetic diseases is the most important outcome. In addition, we will better understand how humans are constructed at the molecular level. As we gain a greater understanding of our genetic makeup, some people are worried that this may lead to greater discrimination. Insurance or health care companies could refuse people who are known to carry genetic abnormalities. Similarly, employers could make their employment decisions based on the genetic makeup of potential employees rather than their past accomplishments. At the family level, genetic information may affect how people choose mates, and whether or not they decide to have children.

CHAPTER 22

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 22.1 A key point is that mRNA is made only when a gene is expressed. In this experiment, mRNA is first isolated and then used to make cDNA, which is fluorescently labeled. The fluorescent spots indicate which genes have been transcribed into mRNA.

FIGURE 22.2 The antibody recognizes the protein of interest that is covalently crosslinked to DNA. The antibody is also attached to heavy beads, which makes it easy to separate it from the rest of the cellular components by centrifugation. This provides an easy way to purify the protein of interest along with its attached DNA.

FIGURE 22.3 Each of these mechanisms alters the structure and possibly the function of proteins, thereby increasing protein diversity.

FIGURE 22.5 The purpose of tandem mass spectrometry is to determine the amino acid sequence of a peptide fragment.

FIGURE 22.6 These homologous genes have similar sequences because they were derived from a common ancestral gene. They are not identical because they have accumulated random mutations after they diverged from each other.

FIGURE 22.7 Yes, you would expect them to be similar because their RNA sequences would also be similar.

End-of-Chapter Questions:

Conceptual Questions

C2. There are two main reasons why the proteome is larger than the genome. The first reason involves the processing of pre-mRNA, a phenomenon that occurs primarily in eukaryotic species. RNA splicing and editing can alter the codon sequence of mRNA, thereby producing alternative forms of proteins that have different amino acid sequences. The second reason for protein diversity is posttranslational modifications. A given protein's structure can be covalently modified by cellular enzymes in many ways. These include proteolytic processing, disulfide bond formation, glycosylation, attachment of lipids, phosphorylation, methylation, and acetylation, to name a few.

C4. Centromeric sequences, origins of replication, telomeric sequences, repetitive sequences, and enhancers. Other examples are possible.

C6. In genetics, the term *similarity* means that two genetic sequences are similar to each other. *Homology* means that two genetic sequences have evolved from a common ancestral sequence. Homologous sequences are similar to each other, but not all short similar sequences are due to homology.

Application and Experimental Questions

E2. In the first dimension (i.e., in the tube gel), proteins migrate to the point in the gel where their net charge is zero. In the second dimension (i.e., the slab gel), proteins are coated with SDS and separated according to their molecular mass.

E4. In tandem mass spectroscopy, the first spectrometer determines the mass of a peptide fragment from a protein of interest. The second spectrometer determines the masses of progressively smaller fragments that are derived from that peptide. Because the masses of each amino acid are known, the molecular masses of these smaller fragments reveal the amino acid sequence of the peptide. With peptide sequence information, it is possible to use the genetic code and produce DNA sequences that could encode such a peptide. More than one sequence is possible due to the degeneracy of the genetic code. These sequences are used as query sequences to search a genomic database. This program will (hopefully) locate a match. The genomic sequence can then be analyzed to determine the entire coding sequence for the protein of interest.

E6. A motif is a sequence that carries out a particular function. There are promoter motifs, enhancer motifs, and amino acid motifs that play functional roles in proteins. For a long genetic sequence, a computer can scan the sequence and identify motifs with great speed and accuracy. The identification of amino acid motifs helps a researcher to understand the function of a particular protein.

E8. In a comparative approach, one uses the sequences of many homologous genes. This method assumes that RNAs of similar function and sequence have a similar structure. For example, computer programs can compare many different 16S rRNA sequences to aid in the prediction of secondary structure.

E10. The three-dimensional structure of a homologous protein must already be solved by X-ray crystallography before one can attempt to predict the three-dimensional structure of a protein of interest based on its amino acid sequence.

E12. The advantages of running a computer program are speed and accuracy. Once the program has been made, and a sequence file has been entered into a computer, the program can analyze long genetic sequences quickly and accurately.

E14. This sequence is within the *lacY* gene of the *lac* operon of *E. coli*, which is described in this chapter. It is nucleotides 801–850.

E16. RNA secondary structure is based on the ability of complementary sequences (i.e., sequences that obey the AU/GC rule) to form a double helix. The program employs a pattern recognition approach. It looks for complementary sequences based on the AU/GC rule.

Questions for Student Discussion/Collaboration

2. This is a very difficult question. In 20 years, we may have enough predictive information so the structure of macromolecules can be predicted from their genetic sequences. If so, it would be better to be a mathematical theoretician with some genetics background. If not, it is probably better to be a biophysicist with some genetics background.

CHAPTER 23

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 23.2 The key feature is that affected offspring have both parents who are unaffected by the disease. The parents are heterozygous carriers.

FIGURE 23.3 The key feature is that affected offspring have an affected parent.

FIGURE 23.4 The key feature is that all of the affected individuals are males. Furthermore, these males all have mothers who were descendents of Queen Victoria.

FIGURE 23.5 *Haplotype* refers to the linkage of alleles or molecular markers along a chromosome.

FIGURE 23.7 In this example, the original mutation that caused the Huntington allele occurred in a germ-line cell or gamete in which the Huntington allele was linked to the G8-C marker.

FIGURE 23.8 The PrP^{Sc} protein may come from eating infected meat or some people have a genetic predisposition that occasionally causes the PrP^{C} protein to convert to the PrP^{Sc} protein.

FIGURE 23.12 Such a mutation would keep the Ras protein in an active state, thereby promoting cancerous growth. The EGF pathway would be turned on.

FIGURE 23.13 The *bcr* gene is expressed in white blood cells. This translocation causes the *abl* gene to be under the control of the *bcr* promoter. The abnormal expression of *abl* in white blood cells causes leukemia.

FIGURE 23.14 E2F will be active all of the time, and this will lead to cancer.

FIGURE 23.16 A checkpoint is a point in the cell cycle in which proteins detect if the cell is in the proper condition to divide. If an abnormality such as DNA damage is detected, the checkpoint proteins will halt the cell cycle.

FIGURE 23.19 An individual with a predisposition for familial breast cancer inherits only one copy of the mutant allele. Therefore, it shows a dominant pattern of inheritance. However, to actually get breast cancer, the other allele must become mutant in somatic cells.

End-of-Chapter Questions:

Conceptual Questions

C2. Recessive X-linked traits are distinguished from the other two by their prevalence in males. Dominant X-linked traits (which are exceedingly rare) are always passed from father to daughter and never from father to son. Autosomal recessive and dominant traits are distinguished primarily by the pattern of transmission from parents to offspring. A person with a dominant trait usually has an affected parent unless it is due to a new mutation or incomplete penetrance is observed. Also, two affected parents can have unaffected children, which is not expected to occur with recessive traits. In the case of rare recessive traits, affected children usually have unaffected parents.

C4. Locus heterogeneity refers to the phenomenon in which a particular type of disease may be caused by mutations in different genes. Hemophilia and thalassemia are two examples. Locus heterogeneity can confound pedigree analysis because alleles in different genes that cause the same disorder may be inherited in different ways. For example, there are autosomal and X-linked forms of hemophilia. If a pedigree contained individuals with both types of alleles, the data would not be consistent with either an autosomal or X-linked pattern of inheritance.

C6. There are lots of possible answers; here are a few. Dwarfism occurs in people and dogs. Breeds like the dachshund and basset hound are types of dwarfism in dogs. There are diabetic people and mice. There are forms of inherited obesity in people and mice. Hip dysplasia is found in people and dogs.

C8. A. Because a person must inherit two defective copies of this gene and it is known to be on chromosome 1, the mode of transmission is autosomal recessive. Both members of this couple must be heterozygous because they have one affected parent (who had to transmit the mutant allele to them) and their phenotypes are unaffected (so they must have received the normal allele from their other parent). Because both parents are heterozygotes, there is a 1/4 chance of producing an affected child (a homozygote) with Gaucher disease. If we let *G* represent the nonmutant allele, and *g* the mutant allele:

♀ \ ♂	*G*	*g*
G	*GG* Normal	*Gg* Normal
g	*Gg* Normal	*gg* Gaucher

B. From this Punnett square, we can also see that there is a 1/4 chance of producing a homozygote with both normal copies of the gene.

C. We need to apply the binomial expansion equation to solve this problem (see Chapter 3 for a description of this equation). In this problem, $n = 5$, $x = 1$, $p = 0.25$, $q = 0.75$. The answer is 0.396, or 39.6%.

C10. The mode of transmission is autosomal recessive. All of the affected individuals do not have affected parents. Also, the disorder is found in both males and females. If it were X-linked recessive, individual III-1 would have to have an affected father, which she does not.

C12. Based on this pedigree, it appears to be an X-linked recessive trait (which it is known to be). All of the affected members are males who had unaffected parents. From this pedigree alone, you could not rule out autosomal recessive, but it is less likely because there are no affected females.

C14. A. The mode of transmission is autosomal recessive. All of the affected individuals do not have affected parents. Also, the disorder is found in both males and females. If it were X-linked recessive, individual III-1 would have to have an affected father, which she does not.

B. If the disorder is autosomal recessive, individuals II-1, II-2, II-6, and II-7 must be heterozygous carriers because they have affected offspring. Because individuals II-2 and II-6 are heterozygotes, one of their parents (I-1 or I-2) must also be a carrier. Because this is rare disease, we would assume that only one of them (I-1or I-2) would be a carrier. Therefore, there is a 50% chance that II-3, II-4, and II-5 are also heterozygotes.

C16. A prion is a protein that behaves like an infectious agent. The infectious form of the prion protein has an abnormal conformation. This abnormal conformation is termed PrP^{Sc}, while the normal conformation of the protein is termed PrP^{C}. An individual can be "infected" with the abnormal conformation of the protein by eating products from another animal that had the disease, or the prion protein may convert spontaneously to the abnormal conformation. A prion protein in the PrP^{Sc} conformation can bind to a prion protein in the PrP^{C} conformation and convert it to the PrP^{Sc} form. An accumulation of prions in the PrP^{Sc} form is what causes the disease symptoms

C18. A. Keep in mind that a conformational change is a stepwise process. It begins in one region of a protein and involves a series of small changes in protein structure. The entire conformational change from PrP^{C} to PrP^{Sc} probably involves many small changes in protein structure that occur in a stepwise manner. Perhaps the conformational change (from PrP^{C} to PrP^{Sc}) begins in the vicinity of position 178 and then proceeds throughout the rest of the protein. If there is a methionine at position 129, the complete conformational change (from PrP^{C} to PrP^{Sc}) can take place. However, perhaps the valine at position 129 somehow blocks one of the steps that are needed to complete the conformational change.

B. Once the PrP^{Sc} conformational change is completed, a PrP^{Sc} protein can bind to another prion protein in the PrP^{C} conformation. Perhaps it begins to convert it (to the PrP^{Sc} conformation) by initiating a small change in protein structure in the vicinity of position 178. The conversion would then proceed in a stepwise manner until the PrP^{Sc} conformation has been achieved. If a valine is at position 129, this could somehow inhibit one of the steps that are needed to complete the conformational change. If an individual had Val-129 in the polypeptide encoded by the second *PrP* gene, half of her or his prion proteins would be less sensitive to conversion by PrP^{Sc} than would be the case for individuals who had Met-129. This would explain why individuals with Val-129 in half of the prion proteins would have disease symptoms that would progress more slowly.

C20. A protooncogene is a normal cellular gene that typically plays a role in cell division. It can be altered by mutation to become an oncogene, thereby causing cancer. At the level of protein function, a protooncogene can become an oncogene by synthesizing too much of a protein or synthesizing the same amount of a protein that is abnormally active.

C22. The predisposition to develop cancer is inherited in a dominant fashion because the heterozygote has the higher predisposition. The mutant allele is actually recessive at the cellular level. But because we have so many cells in our bodies, it is relatively likely that a defective mutation will occur in the single normal gene and lead to a cancerous cell. Some heterozygous individuals may not develop the disease as a matter of chance. They may be lucky and not get a defective mutation in the normal gene or perhaps their immune system is better at destroying cancerous cells once they arise.

C24. A. No, because E2F is inhibited.

B. Yes, because E2F is not inhibited.

C. Yes, because E2F is not inhibited.

D. No, because there is no E2F.

C26. A. True. B. True.

C. False, most cancer cells are caused by mutations that result from environmental mutagens.

D. True.

Application and Experimental Questions

E2. Perhaps the least convincing is the higher incidence of the disease in particular populations. Because populations living in specific geographic locations are exposed to their own unique environment, it is difficult to distinguish genetic versus environmental causes for a particular disease. The most convincing evidence might be the higher incidence of a disease in related individuals and/or the ability to correlate a disease with the presence of a mutant gene. Overall, however, the reliability that a disease has a genetic component should be based on as many observations as possible.

E4. You would probably conclude that it is less likely to have a genetic component. If it were rooted primarily in genetics, it would be likely to be found in the Central American population. Of course, there is a chance that very few or none of the people who migrated to Central America were carriers of the mutant gene, which is somewhat unlikely for a large migrating population. By comparison, one might suspect that an environmental agent that is present in South America but not present in Central America may underlie the disease. Researchers could try to search for this environmental agent (e.g., a pathogenic organism).

E6. Males I-1, II-1, II-4, II-6, III-3, III-8, and IV-5 have a normal copy of the gene. Males II-3, III-2, and IV-4 are hemizygous for an inactive mutant allele. Females III-4, III-6, IV-1, IV-2, and IV-3 have two normal copies of the gene, whereas females I-2, II-2, II-5, III-1, III-5, and III-7 are heterozygous carriers of a mutant allele.

E8. Most inherited forms of cancer are inherited in a dominant manner. This can often be revealed by a pedigree analysis, because affected individuals are much more likely to have affected offspring. Because cancer may be caused by the sequential accumulation of mutations, the correlation between affected parents and affected offspring may be relatively low because the offspring may not accumulate the other postzygotic mutations that are necessary for cancer to occur.

Questions for Student Discussion/Collaboration

2. There isn't a clearly correct answer to this question, but it should stimulate a large amount of discussion.

CHAPTER 24

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 24.2 The mechanisms in parts (a) and (b) involve diffusible morphogens.

FIGURE 24.4 Genes encode proteins that control the changes that cause development to happen. It is largely a hierarchy in which certain sets of genes control other sets of genes.

FIGURE 24.6 The anteroposterior axis is from head to tail. The dorsoventral axis is from back (spine in vertebrates) to front.

FIGURE 24.7 The Bicoid protein promotes the formation of anterior structures in the fruit fly.

FIGURE 24.8 Maternal-effect gene products are first made in nurse cells and are then transported into the oocyte.

FIGURE 24.10 A mutation in a gap gene causes several adjacent segments to be missing, whereas a mutation in a pair-rule gene causes regions in alternating segments to be missing.

FIGURE 24.12 The arrangement of homeotic genes correlates with the anteroposterior axis.

FIGURE 24.15 A cell lineage is a series of cells that are derived from each other via cell division.

FIGURE 24.16 An ortholog is a gene that is homologous to another gene in a different species.

FIGURE 24.18 The expression of the *HoxC-6* gene appears to determine the boundary between the neck and thoracic region in vertebrates.

FIGURE 24.19 The Id protein functions during early stages of embryonic development. It prevents myogenic bHLH proteins from promoting muscle differentiation too soon.

FIGURE 24.22 The correct number of stem cells is needed to promote proper growth and development.

FIGURE 24.24 The whorls would be carpel-stamen-stamen-carpel.

FIGURE 24.26 Site-specific recombination allows a single gene to be rearranged, thereby producing many different types of antibody polypeptides. This greatly increases the diversity of antibodies allowing them to recognize many different types of antigens.

End-of-Chapter Questions:

Conceptual Questions

C2. A. False, the head is anterior to the tail. B. True.

C. False, the feet are ventral to the hips. D. True.

C4. A. True.

B. False, because gradients are also established after fertilization during embryonic development.

C. True.

C6. A. This is a mutation in a pair-rule gene (*runt*).

B. This is a mutation in a gap gene (*knirps*).

C. This is a mutation in a segment-polarity gene (*patched*).

C8. Positional information refers to the phenomenon whereby the spatial locations of morphogens and cell adhesion molecules (CAMs) provide a cell with information regarding its position relative to other cells. In *Drosophila*, the formation of a segmented body pattern relies initially on the spatial location of maternal gene products. These gene products lead to the sequential activation of the segmentation genes. Positional information can come from morphogens that are found within the oocyte, from morphogens secreted from cells during development, and from cell-to-cell contact. Although all three are important, morphogens in the oocyte have the greatest effect on the overall body structure.

C10. The anterior portion of the anteroposterior axis is established by the action of Bicoid. During oogenesis, the mRNA for Bicoid enters the anterior end of the oocyte and is sequestered there to establish an anterior (high) to posterior (low) gradient. Later, when the mRNA is translated, the Bicoid protein in the anterior region establishes a genetic hierarchy that leads to the formation of anterior structures. If Bicoid was not trapped in the anterior end, it is likely that anterior structures would not form.

C12. Maternal-effect gene products influence the formation of the main body axes, including the anteroposterior, dorsoventral, and terminal regions. They are expressed during oogenesis and needed very early in development. Zygotic genes, particularly the three classes of the segmentation genes, are necessary after the axes have been established. The segmentation genes are expressed after fertilization.

C14. The coding sequence of homeotic genes contains a 180-bp consensus sequence known as a homeobox. The protein domain encoded by the homeobox is called a homeodomain. The homeodomain contains three conserved sequences that are folded into α-helical conformations. The arrangement of these α helices promotes the binding of the protein to the major groove of the DNA. Helix III is called the recognition helix because it recognizes a particular nucleotide sequence within the major groove. In this way, homeotic proteins are able to bind to DNA in a sequence-specific manner and thereby activate particular genes.

C16. It would normally be expressed in the three thoracic segments that have legs (T1, T2, and T3).

C18. A. When a mutation inactivates a gap gene, a contiguous section of the larva is missing.

B. When a mutation inactivates a pair-rule gene, some regions that are derived from alternating parasegments are missing.

C. When a mutation inactivates a segment-polarity gene, portions are missing at either the anterior or posterior end of the segments.

C20. Proper development in mammals is likely to require the products of maternal-effect genes that play a key role in initiating embryonic development. The adult body plan is merely an expansion of the embryonic body plan, which is established in the oocyte. Because the starting point for the development of an embryo is the oocyte, this explains why an enucleated oocyte is needed to clone mammals.

C22. Cell differentiation is the specialization of a cell into a particular cell type. In the case of skeletal muscle cells, the bHLH proteins play a key role in the initiation of cell differentiation. When bHLH proteins are activated, they are able to bind to enhancers and activate the expression of many different muscle-specific genes. In this way, myogenic bHLH proteins turn on the expression of many muscle-specific proteins. When these proteins are synthesized, they change the characteristics of the cell into those of a muscle cell. Myogenic bHLH proteins are regulated by dimerization. When a heterodimer forms between a myogenic bHLH protein and an E protein, it activates gene expression. However, when a heterodimer forms between myogenic bHLH proteins and a protein called Id, the heterodimer is unable to bind to DNA. The Id protein is produced during early stages of development and prevents myogenic bHLH proteins from promoting muscle differentiation too soon. At later stages of development, the amount of Id protein falls, and myogenic bHLH proteins can combine with E proteins to induce muscle differentiation.

C24. A totipotent cell is a cell that has the potential to create a complete organism.

A. In humans, a fertilized egg is totipotent, and the cells during the first few embryonic divisions are totipotent. However, after several divisions, embryonic cells lose their totipotency and, instead, are determined to become particular tissues within the body.

B. In plants, many living cells are totipotent.

C. Because yeast are unicellular, one cell is a complete individual. Therefore, yeast cells are totipotent; they can produce new individuals by cell division.

D. Because bacteria are unicellular, one cell is a complete individual. Therefore, bacteria are totipotent; they can produce new individuals by cell division.

C26. Animals begin their development from an egg and then form anteroposterior and dorsoventral axes. The formation of an adult organism is an expansion of the embryonic body plan. Plants grow primarily from two meristems: shoot and root meristems. At the cellular level, plant development is different in that it does not involve cell migration, and most plant cells are totipotent. Animals require the organization within an oocyte to begin development. At the genetic level, however, animal and plant development are similar in that a genetic hierarchy of transcription factors governs pattern formation and cell specialization.

C28. First, gene rearrangement of V, D, and J domains occurs within the light- and heavy-chain genes. Second, within a given B cell, different combinations of light and heavy chains are possible. Third, imprecise fusion may occur between the V, D, and J domains. Fourth, somatic hypermutation may occur due to errors in DNA replication.

C30. One segment, which includes some variable (V) domains and perhaps one or more joining (J) domain, of DNA is removed from the κ light-chain gene. One segment, which may include one or more J domains and the region between the J domain and C domain, is removed during pre-mRNA splicing.

Application and Experimental Questions

E2. *Drosophila* has an advantage in that researchers have identified many more mutant alleles that alter development in specific ways. The hierarchy of gene regulation is particularly well understood in the fruit fly. *C. elegans* has the advantage of simplicity and a complete knowledge of cell fate. This enables researchers to explore how the timing of gene expression is critical to the developmental process.

E4. These results indicate that the gene product is needed from 1 to 3 hours after fertilization for the embryo to develop properly and survive. The gene product is not needed at the other developmental stages that were examined in this experiment (0–1 hours, or 3–6 hours after fertilization).

E6. A. Yes, because Krüppel protein acts as a transcriptional repressor, and its concentration is low in this region anyway.

B. Probably not, because Bicoid protein acts as a transcriptional activator.

C. Probably not, because Hunchback protein acts as a transcriptional activator.

D. Yes, because giant protein acts as a repressor, and its concentration is low in this region anyway.

E8. You could follow the strategy of reverse genetics. Basically, you would create a *HoxD-3* gene knockout. This inactivated *HoxD-3* gene would be introduced into a mouse by the technique of gene replacement described in Chapter 20. By making the appropriate crosses, homozygous mice would be obtained that carry the loss-of-function allele in place of the wild-type *HoxD-3* gene. The phenotypic characteristics of normal mice would then be compared with those of mice that were homozygous for a defective *HoxD-3* gene. This would involve an examination of the skeletal anatomies of mice at various stages of development. If the *HoxD-3* plays a role in development, you might see changes in morphology suggesting anterior transformations. In other words, a certain region of the mouse may have characteristics that are appropriate for more anterior segments.

E10. Geneticists who are interested in mammalian development have used reverse genetics because it has been difficult for them to identify mutations in developmental genes based on phenotypic effects in the embryo. This is because it is difficult to screen a large number of mammalian embryos in search of abnormal ones that carry mutant genes. It is easy to have thousands of flies in a laboratory, but it is not easy to have thousands of mice. Instead, it is easier to clone the normal gene based on its homology to invertebrate genes and then make mutations in vitro. These mutations can be introduced into a mouse to create a gene knockout. This strategy is opposite to that of Mendel, who characterized genes by first identifying phenotypic variants (e.g., tall versus dwarf, green seeds versus yellow seeds, etc.).

Questions for Student Discussion/Collaboration

2. In this problem, the students should try to make a flow diagram that begins with maternal-effect genes, then gap genes, pair-rule genes, and segment polarity genes. These genes then lead to homeotic genes and finally to genes that cause cells to differentiate. It's almost impossible to make an accurate flow diagram because there are so many gene interactions, but it is instructive to think about developmental genetics in this way. It is probably easier to identify mutant phenotypes that affect later stages of development because they are less likely to be lethal. However, modern methods can screen for conditional mutants as described in Solved Problem S3. To identify all of the genes necessary for chicken development, you may begin with early genes but this assumes that you have some way to identify them. If they had been identified, you would then try to identify the genes that they stimulate or repress. This could be done using molecular methods described in Chapters 16, 17, and 19.

CHAPTER 25

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 25.1 A local population is a group of individuals that are more likely to interbreed with each other than with more distant populations.

FIGURE 25.2 Polymorphisms are very common in nearly all natural populations.

FIGURE 25.6 The word *directional* means that selection is favoring a particular phenotype. Selection is moving the population in the direction in which that phenotype will predominate.

FIGURE 25.8 No, because directional selection is eliminating individuals that are sensitive to DDT.

FIGURE 25.9 Stabilizing selection decreases genetic diversity because it eliminates individuals that carry alleles that promote more extreme phenotypes.

FIGURE 25.10 Yes, this selection fosters polymorphism. The fitness values of phenotypes depend on the environment. Some phenotypes are the fittest in one environment, whereas other phenotypes are the fittest in another environment.

FIGURE 25.12 The Hb^S allele is an advantage in the heterozygous condition because it confers resistance to malaria. The heterozygote advantage outweighs the homozygote disadvantage.

FIGURE 25.13 In negative frequency-dependent selection, the rarer phenotype has a higher fitness, which improves its reproductive success.

FIGURE 25.16 Genetic drift tends to have a greater effect in small populations. It can lead to the rapid loss or fixation of an allele.

FIGURE 25.17 At the bottleneck, genetic diversity may be lower because there are fewer individuals. Also, during the time when the bottleneck occurs, genetic drift may promote the loss of certain alleles and the fixation of other alleles, thereby diminishing genetic diversity.

FIGURE 25.18 Inbreeding increases the likelihood of homozygosity, and therefore tends to increase the likelihood that an individual will exhibit a recessive trait. This occurs because an individual can inherit both copies of the same allele from a common ancestor.

FIGURE 25.21 DNA fingerprinting is used for identification and for relationship testing.

End-of-Chapter Questions:

Conceptual Questions

C2. A population is a group of interbreeding individuals. Let's consider a squirrel population in a forested area. Over the course of many generations, several things could happen to this population. A forest fire, for example, could dramatically decrease the number of individuals, thereby causing a bottleneck. This would decrease the genetic diversity of the population. A new predator may enter the region and natural selection may select for the survival of squirrels that are best able to evade the predator. Another possibility is that a group of squirrels within the population may migrate to a new region and found a new squirrel population.

C4. A. Phenotype frequency and genotype frequency.

B. Genotype frequency.

C. Allele frequency.

C6. A. The genotype frequency for the *CF* homozygote is 1/2,500, or 0.004. This equals q^2. The allele frequency is the square root of this value, which equals 0.02. The frequency of the corresponding normal allele equals 1 – 0.02 = 0.98.

B. The frequency for the *CF* homozygote is 0.004; for the unaffected homozygote, $(0.98)^2 = 0.96$; and for the heterozygote, 2(0.98)(0.02), which equals 0.039.

C. If a person is known to be a heterozygous carrier, the chances that this particular person will happen to choose another as a mate is equal to the frequency of heterozygous carriers in the population, which equals 0.039, or 3.9%. The chances that two randomly chosen individuals will choose each other as mates equals $0.039 \times 0.039 = 0.0015$, or 0.15%.

C8. For two alleles, the heterozygote is at a maximum when they are 0.5 each. For three alleles, the two heterozygotes are at a maximum when each allele is 0.33.

C10. Because this is a recessive trait, only the homozygotes for the rolling allele will be able to roll their tongues. If *p* equals the rolling allele and *q* equals the nonrolling allele, the Hardy-Weinberg equation predicts that the frequency of homozygotes who can roll their tongues would be p^2. In this case, $p^2 = (0.6)2 = 0.36$, or 36%.

C12. A. The inbreeding coefficient is calculated using the formula

$$F = \Sigma\ (1/2)^n\ (1 + F_A)$$

In this case there is one common ancestor, I-2. Because we have no prior history on I-2, we assume she is not inbred, which makes $F_A = 0$. The inbreeding loop for IV-3 contains five people, III-4, II-2, I-2, II-5, and III-5. Therefore, $n = 5$.

$$F = (1/2)^5(1 + 0) = 1/32 = 0.031$$

B. Based on the data shown in this pedigree, individual IV-4 is not inbred.

C14. Migration, genetic drift, and natural selection are the main factors that alter allele frequencies within a population. Natural selection acts to eliminate harmful alleles and promote beneficial alleles. Genetic drift involves random changes in allele frequencies that may eventually lead to elimination or fixation of alleles. It is thought to be important in the establishment of neutral alleles in a population. Migration is important because it introduces new alleles into neighboring populations. According to the neutral theory, genetic drift is largely responsible for the variation seen in natural populations.

C16. A random process is one that alters allele frequencies without any regard to whether the changes are beneficial or not. Genetic drift and migration are the two main ways this can occur. An example is the founder effect, when a group of individuals migrates to a new location such as an island. Adaptive processes increase the reproductive success of a species. Natural selection is the adaptive process that tends to eliminate harmful alleles from a population and increase the frequency of beneficial alleles. An example would be the long neck of the giraffe, which enables it to feed in tall trees. At the molecular level, beneficial mutations may alter the coding sequence of a gene and change the structure and function of the protein in a way that is beneficial. For example, the sickle cell anemia allele alters the structure of hemoglobin, and in the heterozygous condition, this decreases the proliferation of the malaria pathogen.

C18. Genetic drift is due to sampling error, and the degree of sampling error depends on the population size. In small populations, the relative proportion of sampling error is much larger. If genetic drift is moving an allele toward fixation, it will take longer in a large population because the degree of sampling error is much smaller.

C20. During the bottleneck effect, allele frequencies are dramatically altered due to genetic drift. In extreme cases, some alleles are lost, whereas others may become fixed at 100%. The overall effect is to decrease genetic diversity within the population. This may make it more difficult for the species to respond in a positive way to changes in the environment. Species that are approaching extinction also face a bottleneck as their numbers decrease. The loss of genetic diversity may make it even more difficult for the species to rebound.

C22. In all cases, these forms of natural selection favor one or more phenotypes because such phenotypes have a reproductive advantage. However, the patterns differ with regard to whether a single phenotype or multiple phenotypes are favored, and whether the phenotype that is favored is in the middle of the phenotypic range or at one or both extremes. Directional selection favors one phenotype at a phenotypic extreme. Over time, natural selection is expected to favor the fixation of alleles that cause these phenotypic characteristics. Disruptive selection favors two or more phenotypic categories. It will lead to a population with a balanced polymorphism for the trait. Examples of balancing selection are heterozygote advantage and negative frequency-dependent selection. These promote a stable polymorphism in a population. Stabilizing selection favors individuals with intermediate phenotypes. It tends to decrease genetic diversity because alleles that favor extreme phenotypes are eliminated.

C24. The intuitive meaning of the mean fitness of a population is the relative likelihood that members of a population will reproduce. If the mean fitness is high, it is likely that an average member will survive and produce offspring. Natural selection increases the mean fitness of a population.

C26. A. True. B. True.

C. False; it causes allele loss or fixation, which results in less diversity.

D. True.

C28. A. Disruptive. Multiple environments favor different phenotypes.

B. Directional. The thicker the fur, the more likely that survival will occur.

C. Stabilizing. Low birth weight is selected against because it results in low survival. Also, very high birth weight is selected against because it could cause problems in delivery, which also could decrease the survival rate.

D. Directional. Sturdy stems and leaves will promote survival in windy climates.

Application and Experimental Questions

E2. ***Answer***: Solved Problem S4 shows how the Hardy-Weinberg equation can be modified to include situations of three or more alleles. In this case:

$$(p + q + r + s)^2 = 1$$

$$p^2 + q^2 + r^2 + s^2 + 2pq + 2qr + 2qs + 2rp + 2rs + 2sp = 1$$

Let $p = C$, $q = c^{ch}$, $r = c^h$, and $s = c$.

A. The frequency of albino rabbits is s^2:

$$s^2 = (0.05) = 0.0025 = 0.25\%$$

B. Himalayan is dominant to albino but recessive to full and chinchilla. Therefore, Himalayan rabbits would be represented by r^2 and by $2rs$:

$$r^2 + 2rs = (0.44)^2 + 2(0.44)(0.05) = 0.24 = 24\%$$

Among 1000 rabbits, about 240 would have a Himalayan coat color.

C. Chinchilla is dominant to Himalayan and albino but recessive to full coat color. Therefore, heterozygotes with chinchilla coat color would be represented by $2qr$ and by $2qs$.

$$2qr + 2qs = 2(0.17)(0.44) + 2(0.17)(0.05) = 0.17, \text{ or } 17\%$$

Among 1000 rabbits, about 170 would be heterozygotes with chinchilla fur.

E4. A. Inuit $M = 0.913$ $N = 0.087$

Navajo $M = 0.917$ $N = 0.083$

Finns $M = 0.673$ $N = 0.327$

Russians $M = 0.619$ $N = 0.381$

Aborigines $M = 0.176$ $N = 0.824$

B. To determine if these populations are in equilibrium, we can use the Hardy-Weinberg equation and calculate the expected number of individuals with each genotype. For example:

Inuit $MM = (0.913)^2 = 83.3$

$MN = 2\,(0.913)(0.087) = 15.9$

$NN = (0.087)^2 = 0.76$

In general, the values agree pretty well with an equilibrium. The same is true for the other four populations.

C. Based on similar allele frequencies, the Inuit and Navajo Indians seem to have interbred as well as the Finns and Russians.

E6. A. $\Delta p_C = m(p_D - p_R)$

With regard to the sickle cell allele:

$$\Delta p_C = (550/10{,}550)(0.1 - 0.01) = 0.0047$$

$$p_C = p_R + \Delta p_C = 0.01 + 0.0047 = 0.0147$$

B. We need to calculate the genotypes separately:

For the 550 migrating individuals,

$Hb^AHb^A = (0.9)^2 = 0.81$, or 81% We expect (0.81)550 = 445.5 individuals to have this genotype

$Hb^AHb^S = 2(0.9)(0.1) = 0.18$ We expect (0.18)550 = 99 heterozygotes

$Hb^SHb^S = (0.1)^2 = 0.01$ We expect (0.01)550 = 5.5 Hb^SHb^S

For the original recipient population,

$Hb^AHb^A = (0.99)^2 = 0.98$ We expect 9801 individuals to have this genotype

$Hb^AHb^S = 2(0.99)(0.01) = 0.0198$ We expect 198 with this genotype

$Hb^SHb^S = (0.01)2 = 0.0001$ We expect 1 with this genotype

To calculate the overall population:

(445.5 + 9801)/10,550 = 0.971 Hb^AHb^A homozygotes

(99 + 198)/10,550 = 0.028 heterozygotes

(5.5 + 1)/10,550 = 0.00062 Hb^SHb^S homozygotes

C. After one round of mating, the allele frequencies in the conglomerate (calculated in part A), should yield the expected genotype frequencies according to the Hardy-Weinberg equilibrium.

Allele frequency of $Hb^S = 0.0147$, so $Hb^A = 0.985$

$Hb^AHb^A = (0.985)^2 = 0.97$

$Hb^AHb^S = 2(0.985)(0.0147) = 0.029$

$Hb^SHb^S = (0.0147)^2 = 0.0002$

E8. A. Probability of fixation = $1/2N$(Assuming equal numbers of males and females contributing to the next generation)

$$\text{Probability of fixation} = 1/2(2{,}000{,}000)$$
$$= 1 \text{ in } 4{,}000{,}000 \text{ chance}$$

B. $\bar{t} = 4N$

Where $\bar{t}$ = the average number of generations to achieve fixation

N = the number of individuals in population, assuming that males and females contribute equally to each succeeding generation

$$\bar{t} = 4(2 \text{ million}) = 8 \text{ million generations}$$

C. If the blue allele had a selective advantage, the value calculated in part A would be slightly larger; there would be a higher chance of allele fixation. The value calculated in part B would be smaller; it would take a shorter period of time to reach fixation.

E10. You could mark snails with a dye and release equal numbers of dark and light snails into dimly lit forested regions and sunny fields. At a later time, recapture the snails and count them. It would be important to have a method of unbiased recapture because the experimenter would have an easier time locating the light snails in a forest and the dark snails in a field. Perhaps one could bait the region with something that the snails like to eat and only collect snails that are at the bait. In addition to this type of experiment, one could also observe predation as it occurs.

E12. Male 2 is the potential father, because he contains the bands that are found in the offspring but are not found in the mother. To calculate the probability, one would have to know the probability of having each of the types of bands that match. In this case, for example, male 2 and the offspring have four bands in common. As a simple calculation, we could eliminate the four bands that the offspring shares with the mother. If the probability of having each paternal band is 1/4, the chances that this person is not the father are $(1/4)^4$.

E14. This percentage is not too high. Based on their genetic relationship, we expect that a father and daughter must share at least 50% of the same bands in a DNA fingerprint. However, the value can be higher than that because the mother and father may have some bands in common, even though they are not genetically related. For example, at one site in the genome, the father may be heterozygous for a 4100-bp and 5200-bp minisatellite, and the mother may also be heterozygous in this same region and have 4100-bp and 4700-bp minisatellites. The father could pass the 5200-bp band to his daughter, and the mother could pass the 4100-bp band. Thus, the daughter would inherit the 4100-bp and 5200-bp bands. This would be a perfect match to both of the father's bands, even though the father transmitted only the 5200-bp band to his daughter. The 4100-bp band matches because the father and mother happened to have a minisatellite in common. Therefore, the 50% estimate of matching bands in a DNA fingerprint based on genetic relationships is a minimum estimate. The value can be higher than that.

Questions for Student Discussion/Collaboration

2. Mutation is responsible for creating new alleles but the rate of new mutations is so low that it cannot explain allele frequencies in this range. Let's call the two alleles *B* and *b* and assume that *B* was the original allele and *b* is a more recent allele that arose as a result of mutation. Three scenarios to explain the allele frequencies are:
   1. The *b* allele is neutral and reached its present frequency by genetic drift. It hasn't reached elimination or fixation yet.
   2. The *b* allele is beneficial and its frequency is increasing due to natural selection. However, there hasn't been enough time to reach fixation.
   3. The *Bb* heterozygote is at a selective advantage leading to a balanced polymorphism.

CHAPTER 26

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 26.1 In most populations (like this one), height follows a continuum.

FIGURE 26.2 About 95.4% are within 2 standard deviations, which means that 4.6 are outside of this range. Half of them, or 2.3%, fall above two standard deviations.

FIGURE 26.3 When alleles are additive, this means they contribute in an incremental way to the outcome of a trait. Having three heavy alleles will make an individual heavier than having two heavy alleles.

FIGURE 26.4 Increases in gene number and more environmental variation tend to cause greater overlaps between different genotypes and the same phenotype.

FIGURE 26.5 These two strains differ with regard to a quantitative trait and they differ in their molecular markers.

FIGURE 26.9 Both natural selection and selective breeding affect allele frequencies due to differences in reproductive success. In the case of natural selection, reproductive success is determined by environmental conditions. For selective breeding, reproductive success is determined by the people who choose the parents for breeding.

FIGURE 26.10 This could be a long list. Examples are weight, disease resistance, food quality, resistance to harsh environmental conditions, and so on.

FIGURE 26.11 A selection limit may be reached because a population has become monomorphic for all of the desirable alleles. A second reason is because the desired effects of artificial selection are balanced by the negative effects on fitness.

End-of-Chapter Questions:

Conceptual Questions

C2. At the molecular level, quantitative traits often exhibit a continuum of phenotypic variation because they are usually influenced by multiple genes that exist as multiple alleles. A large amount of environmental variation will also increase the overlap between genotypes and phenotypes for polygenic traits.

C4. A discontinuous trait is one that falls into discrete categories. Examples include brown eyes versus blue eyes in humans and purple versus white flowers in pea plants. A continuous trait is one that does not fall into discrete categories. Examples include height in humans and fruit weight in tomatoes. Most quantitative traits are continuous; the trait falls within a range of values. The reason why quantitative traits are continuous is because they are usually polygenic and greatly influenced by the environment. As shown in Figure 26.4b, this tends to create ambiguities between genotypes and a continuum of phenotypes.

C6. The top 2.5% is about 2 standard deviation units above the mean. If we take the square root of the variance, the standard deviation would be 22 pounds. To be in the top 2.5% an animal would have to weigh at least 44 pounds heavier than the mean, which equals 606 pounds. To be in the bottom 0.13%, an animal would have to be 3 standard deviations lighter, which would be at least 66 pounds lighter than the mean, or 496 pounds.

C8. There is a positive correlation, but it could have occurred as a matter of chance alone. According to Table 26.2, this value could have occurred through random sampling error. You would need to conduct more experimentation to determine if there is a significant correlation such as examining a greater number of pairs of individuals. If $N = 500$, the correlation would be statistically significant and you would conclude that the correlation did not occur as a matter of random chance. However, you could not conclude cause and effect.

C10. When a correlation coefficient is statistically significant, it means that the association is likely to have occurred for reasons other than random sampling error. It may indicate cause and effect but not necessarily. For example, large parents may have large offspring due to genetics (cause and effect). However, the correlation may be related to the sharing of similar environments rather than cause and effect.

C12. Quantitative trait loci are sites within chromosomes that contain genes that affect a quantitative trait. It is possible for a QTL to contain one gene, or it may contain two or more closely linked genes. QTL mapping, which involves linkage to known molecular markers, is commonly used to determine the locations of QTLs.

C14. If the broad sense heritability equals 1.0, it means that all of the variation in the population is due to genetic variation rather than environmental variation. It does not mean that the environment is unimportant in the outcome of the trait. Under another set of environmental conditions the trait may have turned out quite differently.

C16. When a species is subjected to selective breeding, the breeder is focusing his/her attention on improving one particular trait. In this case, the rose breeder is focused on the size and quality of the flowers. Because the breeder usually selects a small number of individuals (e.g., the ones with best flowers) as the breeding stock for the next generation, this may lead to a decrease in the allelic diversity at other genes. For example, several genes affect flower fragrance. In an unselected population, these genes may exist as "fragrant alleles" and "nonfragrant alleles." After many generations of breeding for large flowers, the fragrant alleles may be lost from the population, just as a matter of random chance. This is a common problem of selective breeding. As you select for an improvement in one trait, you may inadvertently diminish the quality of an unselected trait.

Others have suggested that the lack of fragrance may be related to flower structure and function. Perhaps the amount of energy that a flower uses to make beautiful petals somehow diminishes its capacity to make fragrance.

C18. Broad-sense heritability takes into account all genetic factors that affect the phenotypic variation in a trait. Narrow-sense heritability considers only alleles that behave in an additive fashion. In many cases, the alleles affecting quantitative traits appear to behave additively. More importantly, if a breeder assumes that the heritability of a trait is due to the additive effects of alleles, it is possible to predict the outcome of selective breeding. This is also termed the realized heritability.

C20. A. Because of their good nutrition, you may speculate that they would grow to be taller.

B. If the environment is rather homogeneous, then heritability values tend to be higher because the environment contributes less to the amount of variation in the trait. Therefore, in the commune, the heritability might be higher, because they uniformly practice good nutrition. On the other hand, because the commune is a smaller size than the general population, the amount of genetic variation might be less, so this would make the heritability lower. However, because the problem states that the commune population is large, we would probably assume that the amount of genetic variation is similar to that in the general population. Overall, the best guess would be that the heritability in the commune population is higher because of the uniform nutrition standards.

C. As stated in part B, the amount of variation would probably be similar, because the commune population is large. As a general answer, larger populations tend to have more genetic variation. Therefore, the general population probably has a bit more variation.

C22. A natural population of animals is more likely to have a higher genetic diversity compared to a domesticated population. This is because domesticated populations have been subjected to many generations of selective breeding, which decreases the genetic diversity. Therefore, V_G is likely to be higher for the natural population. The other issue is the environment. It is difficult to say which group would have a more homogeneous environment. In general, natural populations tend to have a more heterogeneous environment, but not always. If the environment is more heterogeneous, this tends to cause more phenotypic variation, which makes V_E higher.

$$\text{Heritability} = V_G/V_T$$
$$= V_G/(V_G + V_E)$$

When V_G is high, this increases heritability. When V_E is high, this decreases heritability. In the natural wolf population, we would expect that V_G would be high. In addition, we would guess that V_E might be high as well (but that is less certain). Nevertheless, if this were the case, the heritability of the wolf population might be similar to the domestic population. This is because the high V_G in the wolf population is balanced by its high V_E. On the other hand, if V_E is not that high in the wolf population, or if it is fairly high in the domestic population, then the wolf population would have a higher heritability for this trait.

Application and Experimental Questions

E2. To calculate the mean, we add the values together and divide by the total number.

$$\text{Mean} = \frac{1.9 + 2(2.4) + 2(2.1) + 3(2.0) + 2(2.2) + 1.7 + 1.8 + 2(2.3) + 1.6}{15}$$

$$\text{Mean} = 2.1$$

The variance is the sum of the squared deviations from the mean divided by $N - 1$. The mean value of 2.1 must be subtracted from each value, and then the square is taken. These 15 values are added together and then divided by 14 (which is $N - 1$).

$$\text{Variance} = \frac{0.85}{14}$$
$$= 0.061$$

The standard deviation is the square root of the variance.

$$\text{Standard deviation} = 0.25$$

E4. When we say that an RFLP is associated with a trait, we mean that a gene that influences a trait is closely linked to an RFLP. At the chromosomal level, the gene of interest is so closely linked to the RFLP that a crossover almost never occurs between them.

Note: Each plant inherits four RFLPs, but it may be homozygous for one or two of them.

Small: 2700 and 4000 (homozygous for both).

Small-medium: 2700 (homozygous), 3000, and 4000; or 2000, 2700, and 4000 (homozygous).

Medium: 2000 and 4000 (homozygous for both); or 2700 and 3000 (homozygous for both); or 2000, 2700, 3000, and 4000.

Medium-large: 2000 (homozygous), 3000, and 4000; or 2000, 2700, and 3000 (homozygous).

Large: 2000 and 3000 (homozygous for both).

E6. Let's assume there is an extensive molecular marker map for the rice genome. We would begin with two strains of rice, one with a high yield and one with a low yield, that greatly differ with regard to the molecular markers that they carry. We would make a cross between these two strains to get F_1 hybrids. We would then backcross the F_1 hybrids to either of the parental strains and then examine hundreds of offspring with regard to their rice yields and molecular markers. In this case, our expected results would be that six different markers in the high-producing strain would be correlated with offspring that produce higher yields. We might get fewer than six bands if some of these genes are closely linked and associate with the same marker. We also might get fewer than six if the two parental strains have the same marker that is associated with one or more of the genes that affect yield.

E8. $h_N^2 = \frac{R}{S}$

$R = \overline{X}_O - \overline{X}$

$S = \overline{X}_P - \overline{X}$

In this problem:

$\overline{X} = 1.01$ (as given in the problem)

$\overline{X}_O = 1.11$ (by calculating the mean for the parents)

$\overline{X}_P = 1.09$ (by calculating the mean for the offspring)

$R = 1.09 - 1.01 = 0.08$

$S = 1.11 - 1.01 = 0.10$

$h_N^2 = 0.08/0.10 = 0.8$ (which is a pretty high heritability value)

E10. We first need to calculate *a* and *b*. In this calculation, *X* represents the height of fathers and *Y* represents the height of sons.

$b = \frac{144}{112} = 1.29$

$a = 69 - (1.29)(68) = -18.7$

For a father who is 70 inches tall,

$$Y = (1.29)(70) + (-18.7) = 71.6$$

The most likely height of the son would be 71.6 inches.

E12. A. After six or seven generations, the selective breeding seems to have reached a plateau. This suggests that the tomato plants have become monomorphic for the alleles that affect tomato weight.

B. One factor may be heterozygote advantage, which is described in Chapter 5 and 25. This partially explains why Martin has obtained tomatoes that are heavier than 1.5 pounds. However, heterozygote advantage is not the whole story; it does not explain why Martin obtained tomatoes that weigh 2 pounds. Even though Mary's and Hector's tomatoes were selected for heavier weight, they may not have all of the "heavy alleles" for each gene that controls weight. For example, let's suppose there are 20 genes that affect weight, with each gene existing in a light and heavy allele. During the early stages of selective breeding, when Mary and Hector picked their 10 plants as seed producers for the next generation, as a matter of random chance, some of these plants may have been homozygous for the light alleles at a few of the 20 genes that control weight. Therefore, just as a matter of chance, they probably "lost" a few of the heavy alleles that affect weight. So, after 12 generations of breeding, they have predominantly heavy alleles but also have light alleles for some of the genes. If we represent heavy alleles with a capital letter, and light alleles with a lowercase letter, Mary's and Hector's strains could be the following:

Mary's strain: *AA BB cc DD EE FF gg hh II JJ KK LL mm NN OO PP QQ RR ss TT*

Hector's strain: *AA bb CC DD EE ff GG HH II jj kk LL MM NN oo PP QQ RR SS TT*

As we see here, Mary's strain is homozygous for the heavy allele at 15 of the genes but carries the light allele at the other 5. Similarly, Hector's strain is homozygous for the heavy allele at 15 genes and carries the light allele at the other 5. It is important to note, however, that the light alleles in Mary's and Hector's strains are not in the same genes. Therefore, when Martin crosses them together, he will initially get

Martin's F_1 offspring: *AA Bb Cc DD EE Ff Gg Hh II Jj Kk LL Mm NN Oo PP QQ RR Ss* TT

If the alleles are additive and contribute equally to the trait, we would expect about the same weight (1.5 pounds), because this hybrid has a total of 10 light alleles. However, if heterozygote advantage is occurring, genes (which were homozygous recessive in Mary's and Hector's strains) will become heterozygous in the F_1 offspring, and this may make the plants healthier and contribute to a higher weight. If Martin's F_1 strain is subjected to selective breeding, the 10 genes that are heterozygous in the F_1 offspring may eventually become homozygous for the heavy allele. This would explain why Martin's tomatoes achieved a weight of 2.0 pounds after five generations of selective breeding.

E14. A. $h_N^2 = \frac{\overline{X}_O - \overline{X}}{\overline{X}_P - \overline{X}}$

$h_N^2 = \frac{269 - 254}{281 - 254} = 0.56$

B. $0.56 = \frac{275 - 254}{\overline{X}_P - 254}$

$\overline{X}_P = 291.5$ lb

Questions for Student Discussion/Collaboration

2. Most traits depend on the influence of many genes. Also, genetic variation is a common phenomenon in most populations. Therefore, most individuals have a variety of alleles that contribute to a given trait. For quantitative traits, some alleles may make the trait bigger and other alleles may make the trait turn out smaller. If a population contains many different genes and alleles that govern a quantitative trait, most individuals will have an intermediate phenotype because they will have inherited some large and some small alleles. Fewer individuals will inherit a predominance of large alleles or a predominance of small alleles. An example of a quantitative trait that does not fit a normal distribution is snail pigmentation. The dark snails and light snails are favored rather than the intermediate colors because they are less susceptible to predation.

CHAPTER 27

Note: the answers to the Comprehension Questions are at the end of the chapter.

Concept Check Questions (in figure legends)

FIGURE 27.1 It illustrates the potential problem of using morphological traits to establish species.

FIGURE 27.2 Cladogenesis is much more common.

FIGURE 27.4 The offspring that are produced from a cross between *G. tetrahit* and either of the other two species will have two sets of chromosomes from one species plus one set from the other. The chromosomes that exist as a single set cannot segregate evenly during meiosis thereby causing sterility. This sterility is why *G. tetrahit* is reproductively isolated from the other two species.

FIGURE 27.5 A clade is a group of species consisting of all of the descendents of the group's most common ancestor.

FIGURE 27.6 A shared derived character is more recent (on an evolutionary time scale) than a primitive character. A shared derived character is shared by a group of organisms but not by a distant common ancestor.

FIGURE 27.7 The idea that the preferred hypothesis is the one that is the simplest.

FIGURE 27.9 The protists are distributed within several supergroups.

FIGURE 27.10 Horizontal gene transfer is the transfer of genetic material to another individual that is not an offspring of that individual. It can occur between different species.

FIGURE 27.11 Orthologs are homologous genes found in different species, whereas paralogs are homologous genes found in the same species.

FIGURE 27.13 A molecule clock allows researchers to put a time scale on their phylogenetic trees.

FIGURE 27.14 In humans, chromosome 2 is a single chromosome but it is divided into two chromosomes in the other species. With regard to chromosome 3, the orangutan has a large inversion that the other species do not have.

FIGURE 27.16 Apoptosis occurs when Gremlin protein does not inhibit BMP4. In the interdigit regions, apoptosis eliminates webbing.

FIGURE 27.17 As the number of *Hox* genes increases, the level of body complexity also tends to increase.

FIGURE 27.18 A *Drosophila* eye is produced because the mouse *Pax6* gene is controlling the genes in the *Drosophila* genome that promote eye development.

End-of-Chapter Questions:

Conceptual Questions

C2. Evolution is unifying because all living organisms on this planet evolved from an interrelated group of common ancestors. At the molecular level, all organisms have a great deal in common. With the exception of a few viruses, they all use DNA as their genetic material. This DNA is found within chromosomes, and the sequence of the DNA is organized into units called genes. Most genes are structural genes that encode the amino acid sequence of polypeptides. Polypeptides fold to form functional units called proteins. At the cellular level, all living organisms also share many similarities. For example, living cells share many of the same basic features including a plasma membrane, ribosomes, enzymatic pathways, and so on. In addition, as discussed in Chapter 6, the mitochondria and chloroplasts of eukaryotic cells are evolutionarily derived from bacterial cells.

C4. Reproductive isolation occurs when two species are unable to mate and produce viable offspring. As discussed in Table 27.1, several prezygotic and postzygotic mechanisms can prevent interspecies matings.

C6. Anagenesis is the evolution of one species into another, whereas cladogenesis is the divergence of one species into two or more species. Of the two, cladogenesis is more prevalent. There may be many reasons why. It is common for an abrupt genetic change such as alloploidy to produce a new species from a preexisting one. Also, migrations of a few members of species into a new region may lead to the formation of a new species in the region (i.e., allopatric speciation).

C8. A. Allopatric B. Sympatric

C. At first, it may involve parapatric speciation with a low level of intermixing. Eventually, when smaller lakes are formed, allopatric speciation will occur.

C10. Reproductive isolation does not really apply to bacteria. Two different bacteria of the same species do not produce gametes that have to fuse to produce an offspring, although bacteria can exchange genetic material (as described in Chapter 9). For this reason, it becomes more difficult to distinguish different species of bacteria. A geneticist would probably divide bacteria into different species based on their cellular traits and the sequences of their DNA. Historically, bacteria were first categorized as different species based on morphological and physiological differences. Later, when genetic tools such as DNA sequencing became available, the previously identified species could be categorized based on genetic sequences. One issue that makes categorization rather difficult is that a species of bacteria can exist as closely related strains that may have a small number of genetic differences.

C12. Line up the sequences where the two Gs are underlined.

TTGCATAGGCATACCGTATGATATCGAAAACTAGAAAAATA
GGGCGATAGCTA

GTATGTTATCGAAAAGTAGCAAAATAGGGCGATAGCTACC
CAGACTACCGGAT

C14. The rate of deleterious and beneficial mutations would probably not be a good molecular clock. Their rate of formation might be relatively constant, but their rate of elimination or fixation would probably be quite variable. These alleles are acted on by natural selection. As environmental conditions change, the degree to which natural selection would favor beneficial alleles and eliminate deleterious alleles would also change. For example, natural selection favors the sickle cell allele in regions where malaria is prevalent but not in other regions. Therefore, the prevalence of this allele does not depend solely on its rate of formation and random genetic drift.

C16. Some regions of a polypeptide are particularly important for the structure or function of a protein. For example, a region of a polypeptide may form the active site of an enzyme. The amino acids that are found within the active site are likely to be precisely located for the binding of the enzyme's substrate and/or for catalysis. Changes in the amino acid sequence of the active site usually have a detrimental effect on the enzyme's functions. Therefore, these types of polypeptide sequences (like those found in active sites) are not likely to change. If they did change, natural selection would tend to prevent the change from being transmitted to future generations. In contrast, other regions of a polypeptide are less important. These other regions would be more tolerant of changes in amino acid sequence and therefore would evolve more rapidly. When comparing related protein sequences, regions that are important for function can often be identified based on less sequence variation.

C18. There are lots of possibilities. A few are listed here.

1. DNA is used as the genetic material.
2. The semiconservative mechanism of DNA replication is the same.
3. The genetic code is fairly universal.
4. Certain genes are found in all forms of life (such as 16S rRNA genes).
5. Gene structure and organization is fairly similar among all forms of life.
6. RNA is transcribed from genes.
7. mRNA is used as a messenger to synthesize polypeptides.

C20. Both theories propose that random mutations occur in populations. The primary difference between the neutral theory and the selectionist theory is the relative contributions of neutral mutations and nonneutral mutations for explaining present-day variation. Both sides agree that natural selection has been an important process in shaping the phenotypes of all species. It is natural selection that favors the traits that allow organisms to survive in their environments. Within each species, however, there is still a great deal of variation (i.e., not all members of the same species are genetically identical). The neutral theory argues that most of this variation is neutral. Random mutations occur that have no effect on the phenotype of the individual, and genetic drift may cause the allele frequency to rise to significant levels. The neutral theory suggests that most (but certainly not all) variation can be explained in this manner. In contrast, the selectionist theory suggests that most of the variation seen in natural populations is due to natural selection. The neutral theory of evolution is sometimes called non-Darwinian evolution because it isn't based on natural selection, which was a central tenet in Darwin's theory.

C22. Natural selection plays a dominant role in the elimination of deleterious mutations and the fixation of beneficial mutations. Genetic drift also would affect the frequencies of deleterious and beneficial alleles, particularly in small populations. Neutral alleles are probably much more frequent than beneficial alleles. Neutral alleles are not acted on by natural selection. Nevertheless, they can become prevalent within a population due to genetic drift. The neutral theory suggests that most of the genetic variation in natural populations is due to the accumulation of neutral mutations. A large amount of data supports this theory, although some geneticists still oppose it.

C24. It would be expressed throughout the bud except in the regions between each finger. Therefore, the cells in these regions undergo apoptosis, which prevents our hands from being webbed. In this way, the cells in between the fingers are eliminated.

Application and Experimental Questions

E2. Perhaps the easiest way to determine allotetraploidy is by the chromosomal examination of closely related species. A researcher could karyotype the chromosomes from many different species and look for homologous chromosomes that have similar banding patterns. This may enable them to identify allotetraploids that contain a diploid set of chromosomes from two different species.

E4. The mutations that have occurred in this sequence are neutral mutations. In all cases, the wobble base has changed, but this change would not affect the amino acid sequence of the encoded polypeptide. Therefore, a reasonable explanation is that the gene has accumulated random neutral mutations over the course of many generations. This observation would be consistent with the neutral theory of evolution. A second explanation would be that one of these two researchers made a few experimental mistakes when determining the sequence of this region.

E6. Inversions do not affect the total amount of genetic material. Usually, inversions do not affect the phenotype of the organism. Therefore, if members of the two populations were to interbreed, the offspring would probably be viable because they would have inherited a normal amount of genetic material from each parent. However, such offspring would be inversion heterozygotes. As described in Chapter 8 (see Figure 8.10), crossing over during meiosis may create chromosomes that have too much or too little genetic material. If these unbalanced chromosomes are passed to the next generation of offspring, the offspring may not survive. For this reason, inversion heterozygotes (that are phenotypically normal) may not be very fertile because many of their offspring will die. Because inversion heterozygotes are less fertile, this would tend to keep the eastern and western populations reproductively isolated. Over time, this would aid in the independent evolution of the two populations and would ultimately promote the evolution of the two populations into separate species.

E8. We would expect the probe to hybridize to the natural *G. tetrahit* and also the artificial *G. tetrahit*, because both of these strains contain two sets of chromosomes from *G. pubescens*. We would expect two bright spots in the in situ experiment. Depending on how closely related *G. pubescens* and *G. speciosa* are, the probe may also hybridize to two sites in the *G. speciosa* genome, but this is difficult to predict a priori. If so, the *G. tetrahit* species would show four spots.

E10. We begin by identifying the pair that is the most closely related. In this case, it is species B and D, because they have the fewest nucleotide differences. We divide 423 by 2 to yield the average number of substitutions, which equals 211.5. The percentage of nucleotide differences between species B and D is:

$$\frac{211.5}{20{,}000} \times 100 = 1.06$$

Species A is the next most closely related species to B and D. The percentage of nucleotide difference between species A compared with B and D is computed as follows:

There are 465 differences between A and D, and 443 differences between A and B. So the average number of differences is:

$$\frac{443 + 465}{2} = 454$$

We divide 454 by 2, which equals 227 to yield the number of nucleotide differences that occur in species A and B or D. The average percentage of nucleotide differences between species A compared with B and D is:

$$\frac{227}{20{,}000} \times 100 = 1.14$$

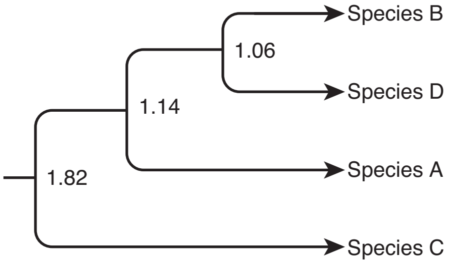
Finally, we consider species C. The percentage of nucleotide difference between species C compared with A, B and D is computed as follows. There are 719 differences between C and A, 744 differences between C and B, and 723 differences between C and D. So the average number of differences is:

$$\frac{719 + 744 + 723}{3} = 729$$

We divide 729 by 2, which equals 364.5, to yield the number of nucleotide differences that occurred in C or A, B, or D. The average percentage of nucleotide differences between species C compared with A, B and D is:

$$\frac{364.5}{20{,}000} \times 100 = 1.82$$

With these data, we can construct the phylogenetic tree shown here:



E12. Possibly, the mouse would have an eye at the tip of its tail!

Questions for Student Discussion/Collaboration

2. The founder effect and allotetraploidy are examples of rapid forms of evolution. In addition, some single gene mutations may have a great influence on phenotype and lead to the rapid evolution of new species by cladogenesis. Some environmental changes may promote the slower accumulation of alleles and alter a species' characteristics more gradually. In this case, it is the accumulation of many phenotypically minor genetic changes that ultimately leads to reproductive isolation. Slow and fast mechanisms of evolution have the common theme that they result in reproductive isolation. This is a prerequisite for the evolution of new species. Fast mechanisms tend to involve small populations and a few number of genetic changes. Slower mechanisms may involve larger populations and involve the accumulation of a large number of genetic changes that each contributes in a small way.

This page intentionally left blank

GLOSSARY

A

A an abbreviation for adenine.

ABC model a model for flower development.

acentric describes a chromosome without a centromere.

acentric fragment a fragment of a chromosome that lacks a centromere.

acquired antibiotic resistance the acquisition of antibiotic resistance because a bacterium has taken up a gene or plasmid from another bacterial strain.

acridine dye a type of chemical mutagen that causes frameshift mutations.

acrocentric a chromosome with the centromere significantly off center, but not at the very end.

activator a transcriptional regulatory protein that increases the rate of transcription.

acutely transforming virus (ACT) a virus that readily transforms normal cells into malignant cells, when grown in a laboratory.

adaptive evolution evolution that is the result of natural selection.

adaptor hypothesis a hypothesis that proposes a tRNA has two functions: recognizing a three-base codon sequence in mRNA and carrying an amino acid that is specific for that codon.

adenine a purine base found in DNA and RNA. It base-pairs with thymine in DNA.

AFLP see *amplified restriction fragment length polymorphism.*

age of onset for alleles that cause genetic diseases, the time of life at which disease symptoms appear.

alkaptonuria a human genetic disorder involving the accumulation of homogentisic acid due to a defect in homogentisic acid oxidase.

alkyltransferase an enzyme that can remove methyl or ethyl groups from guanine bases.

allele an alternative form of a specific gene.

allele frequency the number of copies of a particular allele in a population divided by the total number of all alleles for that gene in the population.

allelic variation genetic variation in a population that involves the occurrence of two or more different alleles for a particular gene.

allodiploid an organism that contains one set of chromosomes from two different species.

allopatric speciation (Greek, *allos*, "other"; Latin, *patria*, "homeland") an evolutionary phenomenon in which speciation occurs when members of a species become geographically separated from the other members.

alloploid an organism that contains chromosomes from two or more different species.

allopoidy the phenomenon in which a cell or organism contains sets of chromosomes from two or more different species.

allopolyploid an organism that contains two (or more) sets of chromosomes from two (or more) species.

allosteric enzyme an enzyme that contains two binding sites: a catalytic site and a regulatory site.

allosteric regulation the phenomenon in which an effector molecule binds to a noncatalytic site on a protein and causes a conformational change that regulates its function.

allosteric site the site on a protein where a small effector molecule binds to regulate the function of the protein.

allotetraploid an organism that contains two sets of chromosomes from two different species.

allozymes two or more enzymes (encoded by the same type of gene) with alterations in their amino acid sequences, which may affect their gel mobilities.

α helix a type of secondary structure found in proteins.

alternative exon an exon that is not always found in mRNA. It is only found in certain types of alternatively spliced mRNAs.

alternative splicing refers to the phenomenon in which a pre-mRNA can be spliced in more than one way.

Ames test a test using strains of a bacterium, *Salmonella typhimurium,* to determine if a substance is a mutagen.

amino acid a building block of polypeptides and proteins. It contains an amino group, a carboxyl group, and a side chain.

aminoacyl site (A site) a site on the ribosome where a charged tRNA initially binds.

aminoacyl tRNA a tRNA molecule that has an amino acid covalently attached to its 3′ end.

aminoacyl-tRNA synthetase an enzyme that catalyzes the attachment of a specific amino acid to the correct tRNA.

2-aminopurine a base analog that acts as a chemical mutagen.

amino terminal end see *amino terminus.*

amino terminus the location of the first amino acid in a polypeptide chain. The amino acid at the amino terminus still retains a free amino group that is not covalently attached to the second amino acid.

amniocentesis a method of obtaining cellular material from a fetus for the purpose of genetic testing.

amplified restriction fragment length polymorphism (AFLP) an RFLP that is amplified via PCR.

anabolic enzyme an enzyme involved in connecting organic molecules to create larger molecules.

anagenesis (Greek, *ana*, "up," and *genesis*, "origin") the evolutionary phenomenon in which a single species is transformed into a different species over the course of many generations.

anaphase the fourth stage of M phase. As anaphase proceeds, half of the chromosomes move to one pole, and the other half move to the other pole.

ancestral character see *primitive character*.

aneuploid not euploid. Refers to a variation in chromosome number such that the total number of chromosomes is not an exact multiple of a set or *n* number.

aneuploidy a cell or organism that is aneuploid.

annealing the process in which two complementary segments of DNA bind to each other.

annotated in files involving genetic sequences, annotation is a description of the known function and features of the sequence, as well as other pertinent information.

anteroposterior axis in animals, the axis that runs from the head (anterior) to the tail or base of the spine (posterior).

anther the structure in flowering plants that gives rise to pollen grains.

anther culture the generation of monoploid plants by cold-shock treatment of anthers.

antibodies proteins produced by the B cells of the immune system that recognize foreign substances (namely, viruses, bacteria, and so forth) and target them for destruction.

antibody microarray a small silica, glass, or plastic slide that is dotted with many different antibodies, which recognize particular amino acid sequences within proteins.

anticipation the phenomenon in which the severity of an inherited disease tends to get worse in future generations.

anticodon a three-nucleotide sequence in tRNA that is complementary to a codon in mRNA.

antigens foreign substances that are recognized by antibodies.

antiparallel an arrangement in a double helix in which one strand is running in the 5′ to 3′ direction, while the other strand is 3′ to 5′.

antisense RNA an RNA strand that is complementary to a strand of mRNA.

antisense strand also called the template strand. It is the strand of DNA that is used as a template for RNA synthesis.

antitermination the function of certain proteins, such as N protein in bacteria, that prevents transcriptional termination.

AP endonuclease a DNA repair enzyme that recognizes a DNA region that is missing a base and makes a cut in the DNA backbone near that site.

apical–basal-patterning gene one of several plant genes that play a role in embryonic development.

apical region in plants, the region that produces the leaves and flowers.

apoptosis programmed cell death.

apurinic site a site in DNA that is missing a purine base.

Archaea also called archaebacteria; one of the three domains of life. Archaea are prokaryotic species. They tend to live in extreme environments and are less common than bacteria.

ARS elements DNA sequences found in yeast that function as origins of replication.

artificial chromosomes cloning vectors that can accommodate large DNA inserts and behave like chromosomes when inside of living cells.

artificial selection see *selective breeding.*

artificial transformation transformation of bacteria that occurs via experimental treatments.

ascus (pl. **asci**) a sac that contains haploid spores of fungi (i.e., yeast or molds).

asexual reproduction a form of reproduction that does not involve the union of gametes; at the cellular level, a preexisting cell divides to produce two new cells.

A site see *aminoacyl site*.

assortative mating breeding in which individuals preferentially mate with each other based on their phenotypes.

AT/GC rule in DNA, the phenomenon in which an adenine base in one strand always hydrogen bonds with a thymine base in the opposite strand, and a guanine base always hydrogen bonds with a cytosine.

ATP-dependent chromatin remodeling see *chromatin remodeling*.

attachment site a site in a host cell chromosome where a virus integrates during site-specific recombination.

attenuation a mechanism of genetic regulation, seen in the *trp* operon, in which a short RNA is made but its synthesis is terminated before RNA polymerase can transcribe the rest of the operon.

attenuator sequence a sequence found in certain operons (e.g., *trp* operon) in bacteria that stops transcription soon after it has begun.

AU-rich element (ARE) a sequence found in many short-lived mRNAs that contains the consensus sequence AUUUA.

automated DNA sequencing the use of fluorescently labeled dideoxyribonucleotides and a fluorescence detector to sequence DNA.

autonomous transposable element a transposable element that contains all of the information necessary for transposition or retroposition to take place.

autopolyploid a polyploid produced within a single species due to nondisjunction.

autoradiography a technique that involves the use of X-ray film to detect the location of radioisotopes as they are found in macromolecules or cells. It is used to detect a particular band on a gel or to map the location of a gene within an intact chromosome.

autoregulatory loop a form of gene regulation in which a protein, such as a splicing factor or a transcription factor, regulates its own expression.

autosomes chromosomes that are not sex chromosomes.

auxotroph a strain that cannot synthesize a particular nutrient and needs that nutrient supplemented in its growth medium or diet.

B

BAC see *bacterial artificial chromosome*.

backbone the portion of a DNA or RNA strand that is composed of the repeated covalent linkage of the phosphates and sugar molecules.

backcross in genetics, this usually refers to a cross of F_1 hybrids to individuals that have genotypes of the parental generation.

Bacteria one of the three domains of living organisms. Also called eubacteria. Bacteria are prokaryotic species.

bacterial artificial chromosome (BAC) a cloning vector that propagates in bacteria and is used to clone large fragments of DNA.

bacteriophages (or **phages**) viruses that infect bacteria.

balanced polymorphism when natural selection favors the maintenance of two or more alleles in a population.

balanced translocation a translocation, such as a reciprocal translocation, in which the total amount of genetic material is normal or nearly normal.

balancing selection a pattern of natural selection that favors the maintenance of two or more alleles. Examples include heterozygote advantage and negative frequency-dependent selection.

band shift assay see *gel retardation assay*.

Barr body a structure in the interphase nuclei of somatic cells of female mammals that is a highly condensed X chromosome.

basal region in plants, the region that produces the roots.

basal transcription in eukaryotes, a low level of transcription via the core promoter. The binding of transcription factors to enhancer elements may increase transcription above the basal level.

basal transcription apparatus the minimum number of proteins needed to transcribe a gene.

base a nitrogen-containing molecule that is a portion of a nucleotide in DNA or RNA. Examples of bases are adenine, thymine, guanine, cytosine, and uracil.

base excision repair (BER) a type of DNA repair in which a modified base is removed from a DNA strand. Following base removal, a short region of the DNA strand is removed, which is then resynthesized using the complementary strand as a template.

base pair the structure in which two nucleotides in opposite strands of DNA hydrogen bond with each other. For example, an AT base pair is a structure in which an adenine-containing nucleotide in one DNA strand hydrogen bonds with a thymine-containing nucleotide in the complementary strand.

base pair mismatch when two bases opposite each other in a double helix do not conform to the AT/GC rule. For example, if A is opposite C, that would be a base mismatch.

base substitution a point mutation in which one base is substituted for another.

basic domain a protein domain containing several basic amino acids, which is often involved in binding to DNA.

Bayesian methods with regard to evolutionary trees, this method asks the question: What is the probability that a particular phylogenetic tree is correct given the observed data and a particular evolutionary model?

B DNA the predominant form of DNA in living cells. It is a right-handed DNA helix with 10 bp per turn.

behavioral trait a trait that involves behavior. An example is the ability to learn a maze.

beneficial mutation a mutation that has a beneficial effect on phenotype.

benign a noncancerous tumor that is not invasive and cannot metastasize.

β sheet a type of secondary structure found in proteins.

bHLH a structure found in transcription factor proteins with a basic domain involved in DNA binding and a helix-loop-helix domain involved in dimerization.

bidirectional in DNA replication, refers to the movement of two different forks outward from a single origin of replication; in gene regulation, refers to the phenomenon that regulatory elements can function in either direction.

bidirectionally the phenomenon in which two replication forks move in opposite directions outward from the origin.

bidirectional replication the phenomenon in which two DNA replication forks emanate in both directions from an origin of replication.

bilateral gynandromorph an animal in which one side is phenotypically male and the other side is female.

bilateralism an animal with left-right symmetry.

bilaterians animals in which the body is organized along an anteroposterior axis that exhibits left-right symmetry.

binary fission the physical process whereby a bacterial cell divides into two daughter cells. During this event, the two daughter cells become divided by the formation of a septum.

binomial expansion equation an equation used to solve genetic problems involving two types of unordered events.

biodegradation the breakdown of a larger molecule into a smaller molecule via cellular enzymes.

bioinformatics the study of biological information. Recently, this term has been associated with the analysis of genetic sequences, using computers and computer programs.

biolistic gene transfer the use of microprojectiles to introduce DNA into plant cells.

biological control the use of microorganisms or products from microorganisms to alleviate plant diseases or damage from undesirable environmental conditions (e.g., frost damage).

biological evolution the accumulation of genetic changes in a species or population over the course of many generations.

biological species concept definition of a species as a group of individuals whose members have the potential to interbreed with one another in nature to produce viable, fertile offspring, but that cannot interbreed successfully with members of other species.

biometric field a field of genetics that involves the statistical study of biological traits.

bioremediation the use of microorganisms to decrease pollutants in the environment.

biotechnology technologies that involve the use of living organisms, or products from living organisms, as a way to benefit humans.

biotransformation the conversion of one molecule into another via cellular enzymes. This term is often used to describe the conversion of a toxic molecule into a nontoxic one.

bivalent a structure in which two pairs of homologous sister chromatids have synapsed (i.e., aligned) with each other.

BLAST (basic local alignment search tool) a computer program that can start with a particular genetic sequence and then locate homologous sequences within a large database.

blending hypothesis of inheritance an early, incorrect hypothesis of heredity. According to this view, the seeds that dictate hereditary traits are able to blend together from generation to generation. The blended traits would then be passed to the next generation.

bottleneck effect a type of genetic drift that occurs when most members of a population are eliminated without any regard to their genetic composition.

box in genetics, a term used to describe a sequence with a specialized function.

branch migration the lateral movement of a Holliday junction.

breakpoint the region where two chromosome pieces break and rejoin with other chromosome pieces.

broad sense heritability heritability that takes into account all genetic factors.

5-bromodeoxyuridine (BrdU) a base analog that can be incorporated into chromosomes during DNA replication.

5-bromouracil a base analog that acts as a chemical mutagen.

C

C an abbreviation for cytosine.

cAMP see *cyclic AMP.*

cAMP response element (CRE) a short DNA sequence found next to certain eukaryotic genes that is recognized by the cAMP response element-binding (CREB) protein.

cancer a disease characterized by uncontrolled cell division.

cancer cell a cell that has lost its normal growth control. Cancer cells are invasive (i.e., they can invade normal tissues) and metastatic (i.e., they can migrate to other parts of the body).

CAP an abbreviation for the catabolite activator protein, a genetic regulatory protein found in bacteria.

capping the covalent attachment of a 7-methylguanosine nucleotide to the 5′ end of mRNA in eukaryotes.

capsid the protein coat of a virus.

CAP site the sequence of DNA that is recognized by CAP.

capsomere a protein subunit of a viral capsid.

carbohydrate organic molecules with the general formula C(H_2O). An example of a simple carbohydrate is the sugar glucose. Large carbohydrates are composed of multiple sugar units.

carboxyl terminus the location of the last amino acid in a polypeptide chain. The amino acid at the carboxyl terminus still retains a free carboxyl group that is not covalently attached to another amino acid.

carcinogen an agent that can cause cancer.

caspases proteolytic enzymes that play a role in apoptosis.

catabolic enzyme an enzyme that is involved in the breakdown of organic molecules into smaller units.

catabolite activator protein see *CAP.*

catabolite repression the phenomenon in which a catabolite (e.g., glucose) represses the expression of certain genes (e.g., the *lac* operon).

catenane interlocked circular molecules.

cDNA see *complementary DNA.*

cDNA library a DNA library made from a collection of cDNAs.

cell adhesion when the surfaces of cells bind to each other or to the extracellular matrix.

cell adhesion molecule (CAM) a molecule (e.g., surface protein or carbohydrate) that plays a role in cell adhesion.

cell culture refers to the growth of cells in a laboratory.

cell cycle in eukaryotic cells, a series of stages through which a cell progresses in order to divide. The phases are G for growth, S for synthesis (of the genetic material), and M for mitosis. There are two G phases, G_1 and G_2.

cell fate the final morphological features that a cell or group of cells will adopt.

cell-free translation system an experimental mixture that can synthesize polypeptides.

cell fusion the process in which individual cells are mixed together and made to fuse with each other.

cell lineage a series of cells that are descended from a cell or group of cells by cell division.

cell lineage diagram in developmental biology, a depiction of a cell and all of its descendants that are produced by cell division.

cell plate the structure that forms between two daughter plant cells that leads to the separation of the cells by the formation of an intervening cell wall.

cellular level with regard to gene expression, a level of observation in which genes affect the traits of cells.

cellular trait a trait that is observed at the cellular level. An example is the shape of a cell.

centiMorgans (cM) (same as a map unit) a unit of map distance obtained from genetic crosses. Named in honor of Thomas Hunt Morgan.

central dogma of genetics the idea that the usual flow of genetic information is from DNA to RNA to polypeptide (protein). In addition, DNA replication serves to copy the information so that it can be transmitted from cell to cell and from parent to offspring.

central region in plants, it is the region that creates the stem. It is the radial pattern of cells in the central region that causes the radial growth observed in plants.

central zone in plants, an area in the meristem where undifferentiated stem cells are always maintained.

centrifugation a method to separate cell organelles and macromolecules in which samples are placed in tubes and spun very rapidly. The rate at which particles move toward the bottom of the tube depends on their densities, sizes, shapes, and the viscosity of the medium.

centrifuge a machine that contains a motor, which causes a rotor holding centrifuge tubes to spin very rapidly.

centrioles in animal cells, a pair of cylindrically shaped structures found at the centrosome.

centromere a segment of eukaryotic chromosomal DNA that provides an attachment site for the kinetochore.

centrosome a cellular structure from which microtubules emanate.

chain termination the stoppage of growth of a DNA strand, RNA strand, or polypeptide sequence.

chaperone a protein that aids in the folding of polypeptides.

character in genetics, this word refers to a general characteristic of an organism.

Chargaff's rule the observation that in DNA, the amount of A equals T, and the amount of G equals C.

charged tRNA a tRNA that has an amino acid attached to its 3′ end by an ester bond.

checkpoint protein a protein that monitors the conditions of DNA and chromosomes and may prevent a cell from progressing through the cell cycle if an abnormality is detected.

chiasma (pl. **chiasmata**) the site where crossing over occurs between two chromosomes. It resembles the Greek letter chi, χ.

chimera an organism composed of cells that are embryonically derived from two different individuals.

ChIP chip assay a form of chromatin immunoprecipitation that utilizes a microarray to determine where in the genome a particular protein binds.

chi square (χ^2) test a commonly used statistical method for determining the goodness of fit. This method can be used to analyze population data in which the members of the population fall into different categories.

chloroplast DNA (cpDNA) the genetic material found within a chloroplast.

chorionic villus sampling a method for obtaining cellular material from a fetus for the purpose of genetic testing.

chromatid following chromosomal replication in eukaryotes, the two copies that remain attached to each other in the form of sister chromatids.

chromatin the association between DNA and proteins that is found within chromosomes.

chromatin immunoprecipitation (ChIP) a method for determining whether proteins bind to a particular region of DNA. This method analyzes DNA-protein interactions as they occur in the chromatin of living cells.

chromatin remodeling a change in chromatin structure that alters the degree of compaction and/or the spacing (or both) of nucleosomes.

chromatography a method of separating different macromolecules and small molecules based on their chemical and physical properties. A sample is dissolved in a liquid solvent and exposed to some type of matrix, such as a gel, a column containing beads, or a thin strip of paper.

chromocenter the central point where polytene chromosomes aggregate.

chromomere a dark band within a polytene chromosome.

chromosome the structures within living cells that contain the genetic material. Genes are physically located within the structure of chromosomes. Biochemically, chromosomes contain a very long segment of DNA, which is the genetic material, and proteins, which are bound to the DNA and provide it with an organized structure.

chromosome map a chart that depicts the linear arrangement of genes along a chromosome.

chromosome painting the use of probes to identify particular regions of chromosomes. The probes are usually assigned a computer-generated color.

chromosome territory in the cell nucleus, each chromosome occupies a nonoverlapping region called a chromosome territory.

chromosome theory of inheritance a theory of Sutton and Boveri that the inheritance patterns of traits can be explained by the transmission patterns of chromosomes during gametogenesis and fertilization.

chromosome walking a common method used in positional cloning in which a mapped gene or RFLP marker provides a starting point to molecularly "walk" toward a gene of interest via overlapping clones.

***cis*-acting element** a sequence of DNA, such as a regulatory element, that exerts a *cis*-effect.

***cis* configuration** when two genetic elements are located on the same chromosome.

***cis*-effect** an effect on gene expression due to genetic sequences that are within the same chromosome and often are immediately adjacent to the gene of interest.

cistron the smallest genetic unit that produces a positive result in a complementation experiment. A cistron is equivalent to a gene.

clade a group of species consisting of all descendents of the group's most common ancestor.

cladistic approach a way to construct a phylogenetic tree, also called a cladogram, by considering the various possible pathways of evolution and then choosing the most plausible tree.

cladogenesis (Greek, *clados*, "branch") during evolution, a form of speciation that involves the division of a species into two or more species.

cladogram a phylogenetic tree that has been constructed using a cladistic approach.

cleavage furrow a constriction that causes the division of two animal cells during cytokinesis.

clonal an adjective to describe a clone. For example, a clonal population of cells is a group of cells that are derived from the same cell.

clone the general meaning of this term is to make many copies of something. In genetics, this term has several meanings: (1) a single cell that has divided to produce a colony of genetically identical cells; (2) an individual that has been produced from a somatic cell of another individual, such as the sheep Dolly; (3) many copies of a DNA fragment that are propagated within a vector or produced by PCR.

closed complex the complex between transcription factors, RNA polymerase, and a promoter before the DNA has denatured to form an open complex.

closed conformation a tightly packed conformation of chromatin that cannot be transcribed.

cluster analysis the analysis of microarray data to determine if certain groups (i.e., clusters) of genes are expressed under the same conditions.

cM an abbreviation for centiMorgans; also see *map unit*.

coactivator a protein that does not bind directly to the DNA, but plays a role in the activation of transcription.

coding strand the strand in DNA that is not used as a template for mRNA synthesis.

codominance a pattern of inheritance in which two alleles are both expressed in the heterozygous condition. For example, a person with the genotype I^AI^B has the blood type AB and expresses both surface antigens A and B.

codon a sequence of three nucleotides in mRNA that functions in translation. A start codon, which usually specifies methionine, initiates translation, and a stop codon terminates translation. The other codons specify the amino acids within a polypeptide sequence according to the genetic code.

codon bias in a given species, the phenomenon in which certain codons are used more frequently than others.

coefficient of inbreeding (F) see *inbreeding coefficient*.

cohesin a protein complex that facilitates the alignment of sister chromatids.

colinearity the correspondence between the sequence of codons in the DNA coding strand and the amino acid sequence of a polypeptide.

colony hybridization a technique in which a probe is used to identify bacterial colonies that contain a hybrid vector with a gene of interest.

combinatorial control the phenomenon common in eukaryotes in which the combination of many factors determines the expression of any given gene.

common ancestor someone who is an ancestor to both of an individual's parents.

comparative genomic hybridization (CGH) a hybridization technique to determine if cells (e.g., cancer cells) have changes in chromosome structure, such as deletions or duplications.

comparative genomics uses information from genome projects to understand the genetic variation between different populations and evolutionary relationships among different species.

competence factors proteins that are needed for bacterial cells to become naturally transformed by extracellular DNA.

competence-stimulating peptide (CSP) a peptide secreted by certain species of bacteria that allow them to become competent for transformation.

competent cells cells that can be transformed by extracellular DNA.

complementary describes sequences in two DNA strands that match each other according to the AT/GC rule. For example, if one strand has the sequence of ATGGCGGATTT, then the complementary strand must be TACCGCCTAAA.

complementary DNA (cDNA) DNA that is made from an RNA template by the action of reverse transcriptase.

complementation a phenomenon in which the presence of two different mutant alleles in the same organism produces a wild-type phenotype. It usually happens because the two mutations are in different genes, so the organism carries one copy of each mutant allele and one copy of each wild-type allele.

complementation test an experimental procedure in which the goal is to determine if two different mutations that affect the same trait are in the same gene or in two different genes.

complete nondisjunction during meiosis or mitosis, when all of the chromosomes fail to disjoin and remain in one of the two daughter cells.

complete transposable element see *autonomous transposable element*.

complex traits characteristics that are determined by several genes and are significantly influenced by environmental factors.

composite transposon a transposon that contains additional genes, such as antibiotic resistance genes, that are not necessary for transposition per se.

computer data file a file (a collection of information) stored by a computer.

computer program a series of operations that can analyze data in a defined way.

concordance in genetics, the degree to which pairs of individuals (e.g., identical twins or fraternal twins) exhibit the same trait.

condensation a change in chromatin structure to become more compact.

condense refers to chromosomes forming a more compact structure

condensin a protein complex that plays a role in the condensation of interphase chromosomes to become metaphase chromosomes.

conditional alleles alleles in which the phenotypic expression depends on the environmental conditions. An example is temperature-sensitive alleles, which affect the phenotype only at a particular temperature.

conditional lethal allele an allele that is lethal, but only under certain environmental conditions.

conditional mutant a mutant whose phenotype depends on the environmental conditions, such as a temperature-sensitive mutant.

conglomerate a population composed of members of an original population plus new members that have migrated from another population.

conjugation a form of genetic transfer between bacteria that involves direct physical interaction between two bacterial cells. One bacterium acts as donor and transfers genetic material to a recipient cell.

conjugation bridge a connection between two bacterial cells that provides a passageway for DNA during conjugation.

conjugative plasmid a plasmid that can be transferred to a recipient cell during conjugation.

consensus sequence the most commonly occurring bases within a sequence element.

conservative model an incorrect model in which both strands of parental DNA remain together following DNA replication.

conservative transposition see *simple transposition*.

constitutive exon an exon that is always found in mRNA following splicing.

constitutive gene a gene that is not regulated and has essentially constant levels of expression over time.

constitutive heterochromatin regions of chromosomes that are always heterochromatic and are permanently transcriptionally inactive.

contig a series of clones that contain overlapping pieces of chromosomal DNA.

continuous traits traits in which the phenotype falls along a continuum.

control element see *regulatory sequence* or *element*.

copy number variation (CNV) variation in the copy number of a gene within a species.

core enzyme the subunits of an enzyme that are needed for catalytic activity, as in the core enzyme of RNA polymerase.

corepressor a small effector molecule that binds to a repressor protein, thereby causing the repressor protein to bind to DNA and inhibit transcription.

core promoter a DNA sequence that is absolutely necessary for transcription to take place. It provides the binding site for general transcription factors and RNA polymerase.

correlation coefficient (*r*) a statistic with a value that ranges between −1 and 1. It describes how two factors vary relative to each other.

cosmid a vector that is a hybrid between a plasmid vector and phage λ. Cosmid DNA can replicate in a cell like a plasmid or be packaged into a protein coat like a phage. Cosmid vectors can accept fragments of DNA that are typically tens of thousands of base pairs in length.

cotransduction the phenomenon in which bacterial transduction transfers a piece of DNA carrying two closely linked genes.

cotransformation the phenomenon in which bacterial transformation transfers a piece of DNA carrying two closely linked genes.

cotranslational events that occur during translation.

coupling in linkage studies, refers to the phenomenon that all of the dominant alleles are on one chromosome and all of the recessive alleles are on the other chromosome.

covariance a statistic that describes the degree of variation between two variables within a group.

cpDNA an abbreviation for chloroplast DNA.

CpG island a group of CG sequences that may be clustered near a promoter region of a gene. The methylation of the cytosine bases usually inhibits transcription.

cross a mating between two distinct individuals. An analysis of their offspring may be conducted to understand how traits are passed from parent to offspring.

cross-fertilization same meaning as *cross*. It requires that the male and female gametes come from separate individuals.

crossing over a physical exchange of chromosome pieces that most commonly occurs during prophase of meiosis I.

C-terminus see *carboxyl terminus*.

cycle threshold method (C_t method) in quantitative PCR, a method of determining the starting amount of DNA based on a threshold level at which the accumulation of fluorescence is significantly greater than the background level.

cyclic AMP (cAMP) in bacteria, a small effector molecule that binds to CAP (catabolite activator protein). In eukaryotes, cAMP functions as a second messenger in a variety of intracellular signaling pathways.

cyclin a type of protein that plays a role in the regulation of the eukaryotic cell cycle.

cyclin-dependent protein kinases (CDKs) enzymes that are regulated by cyclins and can phosphorylate other cellular proteins by covalently attaching a phosphate group.

cytogeneticist a scientist who studies chromosomes under the microscope.

cytogenetic mapping the mapping of genes or genetic sequences using microscopy.

cytogenetics the field of genetics that involves the microscopic examination of chromosomes.

cytokinesis the division of a single cell into two cells. The two nuclei produced in M phase are segregated into separate daughter cells during cytokinesis.

cytological mapping see *cytogenetic mapping.*

cytoplasmic inheritance (also known as *extranuclear inheritance*) the inheritance of genetic material that is not found within the cell nucleus.

cytosine a pyrimidine base found in DNA and RNA. It base-pairs with guanine in DNA.

D

Dam see *DNA adenine methyltransferase.*

Darwinian fitness the relative likelihood that a genotype will contribute to the gene pool of the next generation as compared with other genotypes.

database a computer storage facility that stores many data files such as those containing genetic sequences.

daughter strand in DNA replication, the newly made strand of DNA.

deamination the removal of an amino group from a molecule. For example, the removal of an amino group from cytosine produces uracil.

decoding function the ability of the 16S rRNA to detect when an incorrect tRNA is bound at the A site and prevent elongation until the mispaired tRNA is released from the A site.

decondensed chromosomes that form a less compact structure.

deficiency condition in which a segment of chromosomal material is missing.

degeneracy in genetics, this term means that more that one codon specifies the same amino acid. For example, the codons GGU, GGC, GGA, and GGG all specify the amino acid glycine.

degrees of freedom in a statistical analysis, the number of categories that are independent of each other.

deleterious mutation a mutation that is detrimental with regard to its effect on phenotype.

deletion condition in which a segment of DNA is missing.

deletion mapping the use of strains carrying deletions within a defined region to map a mutation of unknown location.

deme see *local population.*

***de novo* methylation** the methylation of DNA that has not been previously methylated. This is usually a highly regulated event.

deoxyribonucleic acid (DNA) the genetic material. It is a double-stranded structure, with each strand composed of repeating units of deoxyribonucleotides.

deoxyribose the sugar found in DNA.

depurination the removal of a purine base from DNA.

determination the process in which a cell or region of an organism has adopted a particular fate that will occur later in development.

determined cell a cell that is destined to differentiate into a specific cell type.

developmental genetics the area of genetics concerned with the roles of genes in orchestrating the changes that occur during development.

diakinesis the fifth stage of prophase of meiosis I.

diauxic growth the sequential use of two sugars by a bacterium.

dicentric describes a chromosome with two centromeres.

dicentric bridge the region between the two centromeres in a dicentric chromosome.

dicer an endonuclease that makes a cut in double-stranded RNA.

dideoxyribonucleotide (ddNTP) a nucleotide used in DNA sequencing that is missing the 3′—OH group. If a dideoxyribonucleotide is incorporated into a DNA strand, it stops any further growth of the strand.

dideoxy sequencing a method of DNA sequencing that uses dideoxyribonucleotides to terminate the growth of DNA strands.

differential centrifugation a form of centrifugation involving a series of centrifugation steps in which the supernatant or pellet is used in each subsequent centrifugation step.

differentially methylated region (DMR) in the case of imprinting, a site that is methylated during spermatogenesis or oogenesis, but not both.

differentiated cell a cell that has become a specialized type of cell within a multicellular organism.

dihybrid testcross a cross in which an experimenter crosses an individual that is heterozygous for two genes to an individual that is homozygous recessive for the same two genes.

dimeric DNA polymerase a complex of two DNA polymerase proteins that move as a unit during DNA replication.

dioecious in plants, a species in which male and female gametophytes are produced on a single (sporophyte) individual.

diploid an organism or cell that contains two copies of each type of chromosome.

diplotene the fourth stage of prophase of meiosis I.

directionality in DNA and RNA, refers to the 5′ to 3′ arrangement of nucleotides in a strand; in proteins, refers to the linear arrangement of amino acids from the N-terminal to C-terminal ends.

directional selection natural selection that favors an extreme phenotype. This usually leads to the fixation of the favored allele.

direct repeat (DR) short DNA sequences that flank transposable elements in which the DNA sequence is repeated in the same direction.

discontinuous trait a trait in which each offspring can be put into a particular phenotypic category.

discovery-based science experimentation that does not require a preconceived hypothesis. In some cases, the goal is to collect data to be able to formulate a hypothesis.

disequilibrium in population genetics, refers to a population that is not in Hardy-Weinberg equilibrium.

dispersive model an incorrect model for DNA replication in which segments of parental DNA and newly made DNA are interspersed in both strands following the replication process.

disruptive selection natural selection that favors both extremes of a phenotypic category. This results in a balanced polymorphism.

dizygotic twins also known as fraternal twins; twins formed from separate pairs of sperm and egg cells.

DNA the abbreviation for deoxyribonucleic acid.

DNA adenine methyltransferase an enzyme in bacteria that attaches methyl groups to the adenine base in DNA that is found within the sequence GATC.

DnaA box sequence serves as a recognition site for the binding of the DnaA protein, which is involved in the initiation of bacterial DNA replication.

DnaA protein a protein that binds to the dnaA box sequence at the origin of replication in bacteria and initiates DNA replication.

DNA fingerprinting a technology for identifying a particular individual based on the properties of their DNA.

DNA gap repair synthesis the synthesis of DNA in a region where a DNA strand has been previously removed, usually by a DNA repair enzyme or by an enzyme involved in homologous recombination.

DNA gyrase also known as topoisomerase II; an enzyme that introduces negative supercoils into DNA using energy from ATP. Gyrase can also relax positive supercoils when they occur.

DNA helicase an enzyme that separates the two strands of DNA.

DNA library a collection of many hybrid vectors, each vector carrying a particular fragment of DNA from a larger source. For example, each hybrid vector in a DNA library might carry a small segment of chromosomal DNA from a particular species.

DNA ligase an enzyme that catalyzes a covalent bond between two DNA fragments.

DNA methylation the phenomenon in which an enzyme covalently attaches a methyl group (—CH$_3$) to a base (adenine or cytosine) in DNA.

DNA methyltransferase the enzyme that attaches methyl groups to adenine or cytosine bases.

DNA microarray a small silica, glass, or plastic slide that is dotted with many different sequences of DNA, corresponding to short sequences within known genes.

DNA-*N*-glycosylase an enzyme that can recognize an abnormal base and cleave the bond between it and the sugar in the DNA backbone.

DNA polymerase an enzyme that catalyzes the covalent attachment of nucleotides together to form a strand of DNA.

DNA probe in a hybridization experiment, a single-stranded DNA fragment with a base sequence that is complementary to a gene of interest. The DNA probe is labeled to detect a gene of interest.

DNA profiling see *DNA fingerprinting.*

DNA replication the process in which original DNA strands are used as templates for the synthesis of new DNA strands.

DNA replication licensing in eukaryotes, occurs when MCM helicase is bound at an origin, enabling the formation of two replication forks.

DNase an enzyme that cuts the sugar-phosphate backbone in DNA.

DNase I an endonuclease that cleaves DNA.

DNase I footprinting a method for studying protein-DNA interactions in which the binding of a protein to DNA protects the DNA from digestion by DNase I.

DNase I sensitivity the phenomenon in which DNA is in an open conformation and able to be digested by DNase I.

DNA sequencing a method for determining the base sequence in a segment of DNA.

DNA supercoiling the formation of additional coils in DNA due to twisting forces.

DNA translocases a type of motor protein that can move along DNA or cause DNA to move.

DNA uptake signal sequences DNA sequences found in certain species of bacteria that are needed for a DNA fragment to be taken up during transformation.

domain a segment of a protein that has a specific function.

dominant describes an allele that determines the phenotype in the heterozygous condition. For example, if a plant is *Tt* and has a tall phenotype, the *T* (tall) allele is dominant over the *t* (dwarf) allele.

dominant negative mutation a mutation that produces an altered gene product that acts antagonistically to the normal gene product. Shows a dominant pattern of inheritance.

dorsoventral axis in animals, the axis from the spine (dorsal) to the stomach (ventral).

dosage compensation the phenomenon that in species with sex chromosomes, one of the sex chromosomes is altered so that males and females have similar levels of gene expression, even though they do not contain the same complement of sex chromosomes.

double-barrel shotgun sequencing a type of shotgun sequencing in which DNA fragments are randomly sequenced from both ends.

double helix the arrangement in which two strands of DNA (and sometimes RNA) interact with each other to form a double-stranded helical structure.

double-strand break model a model for homologous recombination in which the event that initiates recombination is a double-strand break in one of the double helices.

down promoter mutation a mutation in a promoter that inhibits the rate of transcription.

down regulation genetic regulation that leads to a decrease in gene expression.

duplication a segment of DNA that is repeated more than once within a genome.

dynamic mutation a type of mutation in which a repeated sequence, such as three nucleotides, can readily increase in number from one generation to the next.

E

ecological species concept a species concept in which each species occupies an ecological niche, which is the unique set of habitat resources that a species requires, as well as the species' influence on the environment and other species.

ecotypes genetically distinct populations that are adapted to their local environments.

egg cell also known as an ovum; it is a female gamete that is usually very large and nonmotile.

electrophoresis the migration of ions or molecules in response to an electric field.

electroporation the use of electric current that creates transient pores in the plasma membrane of a cell to allow entry of DNA.

elongation (1) in transcription, the synthesis of RNA using DNA as a template; (2) in translation, the synthesis of a polypeptide using the information within mRNA.

embryogenesis an early stage of animal and plant development leading to the production of an embryo with organized tissue layers and a body plan organization.

embryonic carcinoma cell (EC cell) a type of pluripotent stem cell found in a specific type of human tumor.

embryonic germ cell (EG cell) a type of pluripotent stem cell found in the gonads of the fetus.

embryonic stem cell (ES cell) a type of pluripotent stem cell found in the early blastocyst.

embryo sac in flowering plants, the female gametophyte that contains an egg cell.

emerging viruses viruses that have arisen recently and have the potential to cause widespread disease.

empirical approach a strategy in which experiments are designed to determine quantitative relationships as a way to derive laws that govern biological, chemical, or physical phenomena.

empirical laws laws that are discovered using an empirical (observational) approach.

endonuclease an enzyme that can cut in the middle of a DNA strand.

endopolyploidy in a diploid individual, the phenomenon in which certain cells of the body may be polyploid.

endosperm in flowering plants, the material in the seed, which is $3n$, that nourishes the developing embryo.

endosymbiosis a symbiotic relationship in which the symbiont actually lives inside (*endo*) the larger of the two species.

endosymbiosis theory the theory that the ancient origin of plastids and mitochondria was the result of certain species of bacteria taking up residence within a primordial eukaryotic cell.

enhancer a DNA sequence that functions as a regulatory element. The binding of a regulatory transcription factor to the enhancer increases the level of transcription.

environment the surroundings an organism experiences.

enzyme a protein that functions to accelerate chemical reactions within the cell.

enzyme adaptation the phenomenon in which a particular enzyme appears within a living cell after the cell has been exposed to the substrate for that enzyme.

epigenetic inheritance an inheritance pattern in which a modification to a nuclear gene or chromosome alters gene expression in an organism, but the expression is not changed permanently over the course of many generations.

episome a segment of bacterial DNA that can exist as an F factor and also integrate into the chromosome.

epistasis an inheritance pattern where one gene can mask the phenotypic effects of a different gene.

epitope the structure on the surface of an antigen that is recognized by an antibody.

equilibrium density centrifugation a form of centrifugation in which the particles sediment through the gradient, reaching a position where the density of the particle matches the density of the solution.

error-prone replication a form of DNA replication carried out by lesion-replicating DNA polymerases that results in a high rate of mutation.

E site see *exit site.*

essential gene a gene that is essential for survival.

ethyl methanesulfonate (EMS) a type of chemical mutagen that alkylates bases (i.e., attaches methyl or ethyl groups).

eubacteria see *Bacteria.*

euchromatin DNA that is not highly compacted and may be transcriptionally active.

eukaryotes (Greek, "true nucleus") one of the three domains of life. A defining feature of these organisms is that their cells contain nuclei bounded by cell membranes. Some simple eukaryotic species are single-celled protists and yeast; more complex multicellular species include fungi, plants, and animals.

euploid describes an organism in which the chromosome number is an exact multiple of a chromosome set.

E-value with regard to the BLAST program, the number of times that the match or a better one would be expected to occur purely by random chance in a search of an entire database.

evo-devo see *evolutionary developmental biology.*

evolution see *biological evolution.*

evolutionary developmental biology (evo-devo) a field of biology that focuses on the role of developmental genes in the formation of traits that are important in the evolution of new species.

evolutonary lineage concept a species concept in which species are defined based on the separate evolution of lineages.

excisionase an enzyme that excises a prophage from a host cell's chromosome.

exit site (E site) a site on the ribosome from which an uncharged tRNA exits.

exon a segment of RNA that is contained within the RNA after splicing has occurred. In mRNA, the coding sequence of a polypeptide is contained within the exons.

exon shuffling the phenomenon in which exons are transferred between different genes during evolution. Some researchers believe that transposable elements have played a role in this phenomenon.

exon skipping when an exon is spliced out of a pre-mRNA.

exonuclease an enzyme that digests an RNA or DNA strand from the end.

expression vector a cloning vector that contains a promoter so that the gene of interest is transcribed into RNA when the vector is introduced into a host cell.

expressivity the degree to which a trait is expressed. For example, flowers with deep red color have a high expressivity of the red allele.

extragenic suppressor a mutation in a gene that suppresses the phenotypic effect of a mutation in a different gene.

extranuclear inheritance (also known as *cytoplasmic inheritance*) the inheritance of genetic material that is not found within the nucleus.

ex vivo approach in the case of gene therapy, refers to genetic manipulations that occur outside the body.

F

F see *inbreeding coefficient.*

facultative heterochromatin heterochromatin that is derived from the conversion of euchromatin to heterochromatin.

fate map a diagram that depicts how cell division proceeds in an organism.

feedback inhibition the phenomenon in which the final product of a metabolic pathway inhibits an enzyme that acts early in the pathway.

fertilization the union of gametes (e.g., sperm and egg) to begin the life of a new organism.

F factor a fertility factor found in certain strains of bacteria in addition to their circular chromosome. Strains of bacteria that contain an F factor are designated F^+; strains without F factors are F^-.

F′ factors an F factor that also carries genes derived from the bacterial chromosome.

F_1 generation the offspring produced from a cross of the parental generation.

F_2 generation the offspring produced from a cross of the F_1 generation.

fidelity a term used to describe the accuracy of a process. If there are few mistakes, a process has a high fidelity.

fine structure mapping also known as intragenic mapping; the aim of fine structure mapping is to ascertain the distances between two (or more) different mutations within the same gene.

fitness see *Darwinian fitness.*

flap endonuclease an endonuclease found in eukaryotes that removes RNA or DNA flaps that are generated during DNA replication.

fluctuation test the experimental procedure of Luria and Delbrück in which the number of random mutations in small bacterial cultures fluctuated more than those in one large culture.

fluorescence in situ hybridization (FISH) a form of in situ hybridization in which the probe is fluorescently labeled.

focus with regard to cancer cells, it means a clump of raised cells that grow without regard to contact inhibition.

footprinting see *DNase I footprinting.*

fork see *replication fork.*

forked-line method a method to solve independent assortment problems in which lines are drawn to connect particular genotypes.

forward mutation a mutation that changes the wild-type genotype into some new variation.

founder with regard to genetic diseases, an individual who lived many generations ago and was the person in which a genetic disease originated.

founder effect changes in allele frequencies that occur when a small group of individuals separates from a larger population and establishes a colony in a new location.

fraction (1) following centrifugation, a portion of the liquid contained within a centrifuge tube; (2) following column chromatography, a portion of the liquid that has been eluted from a column.

frameshift mutation a mutation that involves the addition or deletion of nucleotides not in a multiple of three and thereby shifts the reading frame of the codon sequence downstream from the mutation.

frequency distribution a graph that describes the numbers of individuals that are found in each of several phenotypic categories.

functional genomics the study of gene function at the genome level. It involves the study of many genes simultaneously.

functional protein microarray a type of protein microarray that monitors a particular kind of protein function, such as the ability to bind a specific drug.

G

G an abbreviation for guanine.

gain-of-function mutation a mutation that causes a gene to be expressed in an additional place where it is not normally expressed or during a stage of development when it is not normally expressed.

gamete a reproductive cell (usually haploid) that can unite with another reproductive cell to create a zygote. Sperm and egg cells are types of gametes.

gametogenesis the production of gametes (e.g., sperm or egg cells).

gametophyte the haploid generation of plants.

gap gene one category of segmentation genes.

G bands the chromosomal banding pattern that is observed when the chromosomes have been treated with the chemical dye Giemsa.

gel electrophoresis a method that combines chromatography and electrophoresis to separate molecules and macromolecules. Samples are loaded into wells at one end of the gel, and an electric field is applied across the gel that causes charged molecules to migrate from one side of the gel to the other.

gel mobility shift assay see *gel retardation assay.*

gel retardation assay a technique for studying protein-DNA interactions in which the binding of protein to a DNA fragment retards it mobility during gel electrophoresis.

gene a unit of heredity that may influence the outcome of an organism's traits. At the molecular level, a gene contains the information to make a functional product, either RNA or protein.

gene addition the addition of a cloned gene into a site in a chromosome of a living cell.

gene amplification an increase in the copy number of a gene.

gene chip see *DNA microarray*.

gene cloning the production of many copies of a gene using molecular methods such as PCR or the introduction of a gene into a vector that replicates in a host cell.

gene conversion the phenomenon in which one allele is converted to another due to genetic recombination and DNA repair.

gene dosage effect the phenomenon when the number of copies of a gene affects the phenotypic expression of a trait.

gene duplication an increase in the copy number of a gene. Can lead to the evolution of gene families.

gene expression the process in which the information within a gene is accessed, first to synthesize RNA and usually proteins, and eventually to affect the phenotype of the organism.

gene family two or more different genes within a single species that are homologous to each other because they were derived from the same ancestral gene.

gene flow changes in allele frequencies due to migration.

gene interaction when two or more different genes influence the outcome of a single trait.

gene knockin a type of gene addition in which a gene of interest has been added to a particular site in the mouse genome.

gene knockout when both copies of a normal gene have been replaced by an inactive mutant gene.

gene modifier effect when the allele of one gene modifies the phenotypic effect of the allele of a different gene.

gene mutation a relatively small mutation that affects only a single gene.

gene pool the totality of all genes within a particular population.

generalized transduction a form of transduction in which any piece of the bacterial chromosomal DNA can be incorporated into a phage.

general lineage concept a species concept which states that each species is a population of an independently evolving lineage.

general transcription factor (GTF) one of several proteins that are necessary for basal transcription at the core promoter.

gene rearrangement a rearrangement in segments of a gene, as occurs in antibody precursor genes.

gene redundancy the phenomenon in which an inactive gene is compensated for by another gene with a similar function.

gene regulation the phenomenon in which the level of gene expression can vary under different conditions.

gene replacement the swapping of a cloned gene made experimentally with a normal chromosomal gene found in a living cell.

gene therapy the introduction of cloned genes into living cells in an attempt to cure or alleviate disease.

genetically modified organism (GMO) an organism that has received genetic material via recombinant DNA technology.

genetic approach in research, the study of mutant genes that have abnormal function. By studying mutant genes, researchers may better understand normal genes and normal biological processes.

genetic code the correspondence between a codon (i.e., a sequence of three bases in an mRNA molecule) and the functional role that the codon plays during translation. Each codon specifies a particular amino acid or the end of translation.

genetic cross a mating between two individuals and the analysis of their offspring in an attempt to understand how traits are passed from parent to offspring.

genetic drift random changes in allele frequencies due to sampling error.

genetic linkage see *linkage.*

genetic map a chart that describes the relative locations of genes or other DNA segments along a chromosome.

genetic mapping any method used to determine the linear order of genes as they are linked to each other along the same chromosome.

genetic marker any genetic sequence that is used to mark a specific location on a chromosome.

genetic mosaic see *mosaicism.*

genetic polymorphism when two or more alleles occur in population; each allele is found at a frequency of 1% or higher.

genetic recombination (1) the process in which chromosomes are broken and then rejoined to form a novel genetic combination. (2) the process in which alleles are assorted and passed to offspring in combinations that are different from the parents.

genetics the study of heredity.

genetic screening the use of testing methods to determine if an individual is a heterozygous carrier for or has a genetic disease.

genetic testing the analysis of individuals with regard to their genes or gene products. In many cases, the goal is to determine if an individual carries a mutant gene.

genetic transfer the physical transfer of genetic material from one bacterial cell to another.

genetic variation genetic differences among members of the same species or among different species.

genome all of the chromosomes and DNA sequences that an organism or species can possess.

genome database a database that focuses on the genetic sequences and characteristics of a single species.

genome maintenance refers to cellular mechanisms that either prevent mutations from occurring and/or prevent mutant cells from surviving or dividing.

genome-sequencing projects research endeavors that have the ultimate goal of determining the sequence of DNA bases of the entire genome of a given species.

genomic clone a clone made from the cloning of chromosomal DNA.

genomic imprinting a pattern of inheritance that involves a change in a single gene or chromosome during gamete formation. Depending on whether the modification occurs during spermatogenesis or oogenesis, imprinting governs whether an offspring will express a gene that has been inherited from its mother or father.

genomic library a DNA library made from chromosomal DNA fragments.

genomics the molecular analysis of the entire genome of a species.

genotype the genetic composition of an individual, especially in terms of the alleles for particular genes.

genotype-environment association when certain genotypes are preferentially found in particular environments.

genotype-environment interaction when the environmental effects on phenotype differ according to genotype.

genotype frequency the number of individuals with a particular genotype in a population divided by the total number of individuals in the population.

germ cells the gametes (i.e., sperm and egg cells).

germ line a lineage of cells that gives rise to gametes.

germ-line mutation a mutation in a cell of the germ line.

GloFish genetically modified aquarium fish that glow due to the introduction of genes that encode fluorescent proteins.

glucocorticoid receptor a type of steroid receptor that functions as a regulatory transcription factor.

goodness of fit the degree to which the observed data and expected data are similar to each other. If the observed and predicted data are very similar, the goodness of fit is high.

gradualism an evolutionary hypothesis suggesting that each new species evolves continuously over long spans of time. The principal idea is that large phenotypic differences that cause the divergence of species are due to the accumulation of many small genetic changes.

grooves in DNA, the indentations where the atoms of the bases are in contact with the surrounding water. In B DNA, there is a smaller minor groove and a larger major groove.

group I intron a type of intron found in self-splicing RNA that uses free guanosine in its splicing mechanism.

group II intron a type of intron found in self-splicing RNA that uses an adenine nucleotide within the intron itself in its splicing mechanism.

growth factors protein factors that influence cell division.

guanine a purine base found in DNA and RNA. It base-pairs with cytosine in DNA.

gyrase see *DNA gyrase*.

H

haplodiploid system a species, such as certain bees, in which one sex is haploid (e.g., male) and the other sex is diploid (e.g., female).

haploid describes the phenomenon that gametes contain half the genetic material found in somatic cells. For a species that is diploid, a haploid gamete contains a single set of chromosomes.

haploinsufficiency the phenomenon in which a person has only a single functional copy of a gene, and that single functional copy does not produce a normal phenotype. Shows a dominant pattern of inheritance.

haplotype the linkage of particular alleles or molecular markers along a single chromosome.

haplotype association study a study in which disease-causing alleles are identified due to their linkage to particular markers along a chromosome.

Hardy-Weinberg equation $p^2 + 2pq + q^2 = 1$.

Hardy-Weinberg equilibrium the phenomenon by which under certain conditions, allele frequencies are maintained in a stable condition and genotypes can be predicted according to the Hardy-Weinberg equation.

helicase see *DNA helicase*.

helix–loop–helix domain a domain found in transcription factors that enables them to dimerize.

helix-turn-helix motif a structure found in transcription factor proteins that promotes binding to the major groove of DNA.

hemizygous describes the single copy of an X-linked gene in the male. A male mammal is said to be hemizygous for X-linked genes.

heritability the amount of phenotypic variation within a particular group of individuals that is due to genetic factors.

heterochromatin highly compacted DNA. It is usually transcriptionally inactive.

heterodimer when two polypeptides encoded by different genes bind to each other to form a dimer.

heteroduplex a double-stranded region of DNA that contains one or more base mismatches.

heterogametic sex in species with two types of sex chromosomes, the heterogametic sex produces two types of gametes. For example, in mammals, the male is the heterogametic sex, because a sperm can contain either an X or a Y chromosome.

heterogamous describes a species that produces two morphologically different types of gametes (i.e., sperm and eggs).

heterogeneity see *locus heterogeneity*.

heterokaryon a cell produced from cell fusion that contains two separate nuclei.

heteroplasmy when a cell contains variation in a particular type of organelle. For example, a plant cell could contain some chloroplasts that make chlorophyll and other chloroplasts that do not.

heterosis the phenomenon in which hybrids display traits superior to either corresponding parental strain. Heterosis is usually different from overdominance, because the hybrid may be heterozygous for many genes, not just a single gene, and because the superior phenotype may be due to the masking of deleterious recessive alleles.

heterozygote an individual who is heterozygous.

heterozygote advantage a pattern of inheritance in which a heterozygote has a reproductive advantage compared to either of the corresponding homozygotes.

heterozygous describes a diploid individual that has different copies (i.e., two different alleles) of the same gene.

Hfr **strain** (for **high frequency of recombination**) a bacterial strain in which an F factor has become integrated into the bacterial chromosome. During conjugation, an *Hfr* strain can transfer segments of the bacterial chromosome.

hierarchical shotgun sequencing a genome sequencing strategy in which small DNA fragments are mapped prior to DNA sequencing.

highly repetitive sequences sequences that are found tens of thousands or even millions of times throughout the genome.

high stringency refers to highly selective hybridization conditions that promote the binding of DNA or RNA fragments that are perfect or almost perfect matches.

high-throughput sequencing DNA sequencing technologies that can process a large number of samples and thereby produce a large amount of DNA sequence information in a relatively short period of time.

histone acetyltransferase an enzyme that attaches acetyl groups to the amino terminal tails of histone proteins.

histone code hypothesis the hypothesis that the pattern of histone modification acts much like a language or code in specifying alterations in chromatin structure.

histone deacetylase an enzyme that removes acetyl groups from the amino terminal tails of histone proteins.

histone variants histones that have amino acid sequences that are slightly different from the core histones and play specialized roles in chromatin structure and function.

histone proteins a group of proteins involved in forming the nucleosome structure of eukaryotic chromatin.

holandric gene a gene on the Y chromosome.

Holliday junction a site where an unresolved crossover has occurred between two homologous chromosomes.

Holliday model a model to explain the molecular mechanism of homologous recombination.

holoenzyme an enzyme containing all of its subunits, as in the holoenzyme of RNA polymerase that has σ factor along with the core enzyme.

homeobox a 180-bp consensus sequence found in homeotic genes.

homeodomain the protein domain encoded by the homeobox. The homeodomain promotes the binding of the protein to the DNA.

homeologous describes the homologous chromosomes from closely related species.

homeotic an adjective that was originally used to describe mutants in which one body part is replaced by another.

homeotic gene a gene that functions in governing the developmental fate of a particular region of the body.

homoallelic two or more alleles in different organisms that are due to mutations at exactly the same base within a gene.

homodimer when two polypeptides encoded by the same gene bind to each other to form a dimer.

homogametic sex in species with two types of sex chromosomes, the homogametic sex produces only one type of gamete. For example, in mammals, the female is the homogametic sex, because an egg can only contain an X chromosome.

homolog one of the chromosomes in a pair of homologous chromosomes.

homologous in the case of genes, this term describes two genes that are derived from the same ancestral gene. Homologous genes have similar DNA sequences. In the case of chromosomes, the two homologs of a chromosome pair are said to be homologous to each other.

homologous recombination the exchange of DNA segments between homologous chromosomes.

homologous recombination repair (HRR) also called homology-directed repair, occurs when the DNA strands from a sister chromatid are used to repair a lesion in the other sister chromatid.

homology structures that are similar to each other because they evolved from a common ancestor.

homology-base modeling the modeling of a three-dimensional structure of a molecule based on its homology to another molecule whose structure is already known.

homozygous describes a diploid individual that has two identical alleles of a particular gene.

horizontal gene transfer the transfer of genes from one organism to another organism that is not its offspring.

host cell a cell that is infected with a virus or bacterium.

host range the spectrum of host species that a virus or other pathogen can infect.

hot spots sites within a gene that are more likely to be mutated than other locations.

housekeeping gene a gene that encodes a protein required in most cells of a multicellular organism.

Hox **complexes** a group of several *Hox* genes located in a particular chromosomal region.

Hox **genes** mammalian genes that play a role in development. They are homologous to homeotic genes found in *Drosophila*.

Human Genome Project a worldwide collaborative project that provided a detailed map of the human genome and obtained a complete DNA sequence of the human genome.

human immunodeficiency virus (HIV) an emerging virus that infects humans and other primates and causes acquired immunodeficiency syndrome (AIDS).

hybrid (1) an offspring obtained from a hybridization experiment; (2) a cell produced from a cell fusion experiment in which the two separate nuclei have fused to make a single nucleus.

hybrid cell a cell produced from the fusion of two different cells, usually from two different species.

hybrid dysgenesis a syndrome involving defective *Drosophila* offspring, due to the phenomenon that P elements can transpose freely.

hybridization (1) the mating of two organisms of the same species with different characteristics; (2) the phenomenon in which two single-stranded molecules renature together to form a hybrid molecule.

hybrid vector a cloning vector that contains an insert of foreign DNA.

hybrid vigor see *heterosis*.

hybrid zones during parapatric speciation, places where two populations can interbreed.

hypothesis testing using statistical tests to determine if the data from experimentation are consistent with a hypothesis.

I

ICR see *imprinting control region*.

illegitimate recombination see *nonhomologous recombination*.

immunoglobulin (IgG) see *antibodies*.

immunoprecipitation the use of antibodies to cause other molecules, such as proteins, to precipitate, which allows them to be collected by centrifugation.

imprinting see *genomic imprinting*.

imprinting control region (ICR) a DNA region that is differentially methylated and plays a role in genomic imprinting.

inborn error of metabolism a genetic disease that involves a defect in a metabolic enzyme.

inbreeding the practice of mating between genetically related individuals.

inbreeding coefficient (*F*) the probability that two alleles in a particular individual will be identical for a given gene because both copies are due to descent from a common ancestor.

inbreeding depression the phenomenon in which inbreeding produces homozygotes that are less fit, thereby decreasing the reproductive success of a population.

incomplete dominance a pattern of inheritance in which a heterozygote that carries two different alleles exhibits a phenotype that is intermediate to the corresponding homozygous individuals. For example, an *Rr* heterozygote may be pink, whereas the *RR* and *rr* homozygotes are red and white, respectively.

incomplete penetrance a pattern of inheritance in which a dominant allele does not always control the phenotype of the individual.

incomplete transposable element see *nonautonomous element*.

induced a gene that has been transcriptionally activated by an inducer.

induced mutation a mutation caused by environmental agents.

inducer a small effector molecule that binds to a genetic regulatory protein and thereby increases the rate of transcription.

inducible gene a gene that is regulated by an inducer, which is a small effector molecule that causes transcription to increase.

induction (1) the effects of an inducer on increasing the transcription of a gene; (2) the process by which a cell or group of cells governs the developmental fate of neighboring cells.

infective particle genetic material found within the cytoplasm of eukaryotic cells that differs from the genetic material normally found in cell organelles.

ingroup in cladistics, a group of species in which a researcher is interested.

inhibitor a small effector molecule that binds to an activator protein, causing the protein to be released from the DNA, thereby inhibiting transcription.

initiation (1) in transcription, the stage that involves the initial binding of RNA polymerase to the promoter in order to begin RNA synthesis; (2) in translation, the formation of a complex between mRNA, the initiator tRNA, and the ribosomal subunits.

initiator tRNA during translation, the tRNA that binds to the start codon.

insertion sequences the simplest transposable elements. They are commonly found in bacteria.

in situ **hybridization** a technique used to cytologically map the locations of genes or other DNA sequences within large eukaryotic chromosomes. In this method, a complementary probe is used to detect the location of a gene within a set of chromosomes.

insulator a segment of DNA that insulates a gene from the effects of nearby regulatory elements such as enhancers.

integrase an enzyme that functions in the integration of viral DNA or retroelements into a chromosome.

interference see *positive interference*.

intergenic mapping a form of mapping that determines the distance between different genes.

intergenic region in a chromosome, a region of DNA that lies between two different genes.

intergenic suppressor a suppressor mutation that is in a different gene from the gene that contains the first mutation.

internal nuclear matrix a network of irregular protein fibers and other proteins that is connected to the nuclear lamina and fills the interior of the nucleus.

interphase the series of phases G_1, S, and G_2, during which a cell spends most of its life.

interrupted mating a method used in conjugation experiments in which the length of time that the bacteria spend conjugating is stopped by a blender treatment or other type of harsh agitation.

interstitial deletion when an internal segment is lost from a linear chromosome.

intervening sequence also known as an *intron*. A segment of RNA that is removed during RNA splicing.

intragenic mapping see *fine structure mapping*.

intragenic suppressor a suppressor mutation that is within the same gene as the first mutation that it suppresses.

intrinsic termination transcriptional termination that does not require the function of the rho protein.

intron intervening sequences that are found between exons. Introns are spliced out of the RNA prior to translation.

invasive refers to a tumor that can invade surrounding tissue.

inversion a change in the orientation of genetic material along a chromosome such that a segment is flipped in the reverse direction.

inversion heterozygote a diploid individual that carries one normal chromosome and a homologous chromosome with an inversion.

inversion loop the loop structure that is formed when the homologous chromosomes of an inversion heterozygote attempt to align themselves (i.e., synapse) during meiosis.

inverted repeats DNA sequences found in transposable elements that are identical (or very similar) but run in opposite directions.

in vitro fertilization (IVF) in the case of humans, the fertilization of an egg outside of a female's body.

iron regulatory protein (IRP) a translational regulatory protein that recognizes iron response elements that are found in specific mRNAs. It may inhibit translation or stabilize the mRNA.

iron response element (IRE) an RNA sequence that is recognized by the iron regulatory protein.

isoacceptor tRNAs two different tRNAs that can recognize the same codon.

isoelectric focusing a form of gel electrophoresis in which a protein migrates to the point in the gel where its net charge is zero.

isogamous describes a species that makes morphologically similar gametes.

isolating mechanism a mechanism that favors reproductive isolation. These can be prezygotic or postzygotic.

K

karyotype a photographic representation of all the chromosomes within a cell. It reveals how many chromosomes are found within an actively dividing somatic cell.

kinetochore a group of cellular proteins that attach to the centromere during meiosis and mitosis.

knockin see *gene knockin.*

knockout see *gene knockout.*

Kozak's rules a set of rules that describes the most favorable types of bases that flank a eukaryotic start codon.

L

lac repressor a protein that binds to the operator site of the *lac* operon and inhibits transcription.

lagging strand a strand during DNA replication that is synthesized as short Okazaki fragments in the direction away from the replication fork.

lariat an excised intron structure composed of a circle and a tail.

latent with regard to viruses, a virus that exists in a dormant state for a long period of time before new virus particles are made.

law of independent assortment see *Mendel's law of independent assortment.*

law of segregation see *Mendel's law of segregation.*

leading strand a strand during DNA replication that is synthesized continuously toward the replication fork.

left-right axis in bilaterians, the axis that determines left-right symmetry.

leptotene the first stage of prophase of meiosis I.

lesion-replicating polymerase a type of DNA polymerase that can replicate over a DNA region that contains an abnormal structure (i.e., a lesion).

lethal allele an allele that may cause the death of an organism.

lethal mutation a mutation that produces a lethal allele that causes the death of an organism.

library see *DNA library.*

ligase see *DNA ligase.*

LINEs in mammals, long interspersed elements that are usually 1 to 5 kbp in length and found in a few thousand to several hundred thousand copies per genome.

linkage the occurrence of two or more genes along the same chromosome.

linkage disequilibrium when alleles and molecular markers are associated with each other at a frequency that is significantly higher than expected by random chance.

linkage group a group of genes that are linked together because they are found on the same chromosome.

linkage mapping the mapping of genes or other genetic sequences along a chromosome by analyzing the outcome of crosses.

lipid a general name given to an organic molecule that is insoluble in water. Cell membranes contain a large amount of lipid.

liposome a vesicle that is surrounded by a phospholipid bilayer.

local population a segment of a population that is slightly isolated. Members of a local population are more likely to breed with each other than with members that are outside of the local population.

locus (pl. **loci**) the physical location of a gene within a chromosome.

locus control region (LCR) a segment of DNA that is involved in the regulation of chromatin opening and closing.

locus heterogeneity the phenomenon in which a particular type of disease or trait may be caused by mutations in two or more different genes.

lod score method a method that analyzes pooled data from a large number of pedigrees or crosses to determine the probability that two genetic markers exhibit a certain degree of linkage. A lod score value ≥ 3 is usually accepted as strong evidence that two markers are linked.

LOH see *loss of heterozygosity.*

long interspersed elements. See LINEs

long terminal repeats (LTRs) sequences containing many short segments that are tandemly repeated. They are found in retroviruses and viral-like retroelements.

loop domain a segment of chromosomal DNA that is anchored by proteins, so it forms a loop.

loss-of-function allele an allele of a gene that encodes an RNA or protein that is nonfunctional or compromised in function.

loss-of-function mutation a change in a genetic sequence that creates a loss-of-function allele.

loss of heterozygosity (LOH) when a heterozygous somatic cell incurs a genetic change that inactivates the single functional allele.

low stringency refers to hybridization conditions in which DNA or RNA fragments that have some mismatches are still able to recognize and bind to each other.

LTR retrotransposons a type retrotransposon that is derived from a virus and has long terminal repeats.

LTRs see *long terminal repeats.*

Lyon hypothesis a hypothesis to explain the pattern of X inactivation seen in mammals. Initially, both X chromosomes are active. However, at an early stage of embryonic development, one of the two X chromosomes is randomly inactivated in each somatic cell.

lysis cell breakage.

lysogenic cycle a type of growth cycle for a phage in which the phage integrates its genetic material into the chromosome of the bacterium. This integrated phage DNA can exist in a dormant state for a long time, during which no new bacteriophages are made.

lysogeny when a virus has integrated its DNA into a host cell's chromosome and exists in a dormant state.

lytic cycle a type of growth cycle for a phage in which the phage directs the synthesis of many copies of the phage genetic material and coat proteins. These components then assemble to make new phages. When synthesis and assembly is completed, the bacterial host cell is lysed, and the newly made phages are released into the environment.

M

macroevolution evolutionary changes at or above the species level involving relatively large changes in form and function that are sufficient to produce new species and higher taxa.

macromolecule a large organic molecule composed of smaller building blocks. Examples include DNA, RNA, proteins, and large carbohydrates.

MADS box a domain found in several transcription factors that play a role in plant development.

maintenance methylation the methylation of hemimethylated DNA following DNA replication.

major groove a wide indentation in the DNA double helix in which the bases have access to water.

malignant describes a tumor composed of cancerous cells.

mammalian cloning experimentally, this refers to the use of somatic cell nuclei and enucleated eggs to create a clone of a mammal.

map distance the relative distance between sites (e.g., genes) along a single chromosome.

mapping the experimental process of determining the relative locations of genes or other segments of DNA along individual chromosomes.

mapping functions approaches to determine the quantitative relationship between recombination frequencies and physical distances along chromosomes.

map unit (mu) a unit of map distance obtained from genetic crosses. One map unit is equivalent to 1% recombinant offspring in a testcross.

mass spectrometry a technique to accurately measure the mass of molecule, such as a peptide fragment.

maternal effect an inheritance pattern for certain nuclear genes in which the genotype of the mother directly determines the phenotypic traits of her offspring.

maternal inheritance inheritance of DNA that occurs through the cytoplasm of the egg.

matrix-attachment region (MAR) a site in the chromosomal DNA that is anchored to the nuclear matrix or scaffold.

maturase a protein that enhances the rate of splicing of Group I and II introns.

maximum likelihood a strategy for choosing the best phylogenetic tree based on assumptions about the most likely ways that DNA incurs mutations.

MCM helicase a group of eukaryotic proteins needed to complete a process called DNA replication licensing, which is necessary for the formation of two replication forks at an origin of replication.

mean ($\overline{X}$) the sum of all the values in a group divided by the number of individuals in the group.

mean fitness of the population ($\overline{W}$) the average fitness of a population that is calculated by considering the frequencies and fitness values for all genotypes.

mediator a protein complex that interacts with RNA polymerase II and various regulatory transcription factors. Depending on its interactions with regulatory transcription factors, mediator may stimulate or inhibit RNA polymerase II.

meiosis a form of nuclear division in which the sorting process results in the production of haploid cells from a diploid cell.

meiotic nondisjunction the event in which chromosomes do not segregate equally during meiosis.

Mendelian inheritance the common pattern of inheritance observed by Mendel, which involves the transmission of eukaryotic genes that are located on the chromosomes found within the cell nucleus.

Mendel's law of independent assortment two different genes randomly assort their alleles during gamete formation (if they are not linked).

Mendel's law of segregation the two copies of a gene segregate from each other during transmission from parent to offspring.

meristem in plants, an organized group of actively dividing cells.

meristic traits traits that can be counted and expressed in whole numbers.

merozygote a partial diploid strain of bacteria containing F' factor genes.

messenger RNA (mRNA) a type of RNA that contains the information for the synthesis of a polypeptide.

metacentric describes a chromosome with the centromere in the middle.

metaphase the third stage of M phase. The chromosomes align along the center of the spindle apparatus.

metaphase plate the plane at which chromosomes align during metaphase.

metastatic describes cancer cells that migrate to other parts of the body.

methylation see *DNA methylation.*

methyl-CpG-binding protein a protein that binds to a CpG island when it is methylated.

methyl-directed mismatch repair a DNA repair system that detects a mismatch and specifically removes the segment from the newly made daughter strand.

methyltransferase see *DNA methyltransferase.*

microarray see *DNA microarray*, *protein microarray*, or *antibody microarray*.

microevolution changes in the gene pool with regard to particular alleles that occur over the course of many generations.

microinjection the use of microscopic-sized needles or pipettes to inject a substance, such as DNA, into cells.

microRNAs small RNA molecules that can inhibit the expression of specific mRNAs via RNA interference.

microsatellite short tandem repeats (typically a couple hundred base pairs in length) that are interspersed throughout a genome and are quite variable in length among different individuals. They can be amplified by PCR.

microscopy the use of a microscope to view cells or subcellular structures.

microtubule-organizing center (MTOC) a site in a cell where microtubules begin to grow.

minichromosomes a structure formed from many copies of circular DNA molecules.

minimal medium a type of growth medium for microorganisms that contains a mixture of reagents that are required for growth; nothing additional has been added.

minisatellite a repetitive sequence that was formerly used in DNA fingerprinting. Its use has been largely superseded by smaller repetitive sequences called microsatellites.

minor groove a narrow indentation in the DNA double helix in which the bases have access to water.

minute a unit of measure in bacterial conjugation experiments. This unit refers to the relative time it takes for genes to first enter a recipient strain during conjugation.

mismatch repair system see *methyl-directed mismatch repair*.

missense mutation a base substitution that leads to a change in the amino acid sequence of the encoded polypeptide.

mitochondrial DNA the DNA found within mitochondria.

mitosis a type of nuclear division into two nuclei, such that each daughter cell receives the same complement of chromosomes.

mitotic nondisjunction an event in which chromosomes do not segregate equally during mitosis.

mitotic recombination crossing over that occurs during mitosis.

mitotic spindle apparatus (also known as the *mitotic spindle*) the structure that organizes and separates the chromosomes during M phase of the eukaryotic cell cycle.

model organism an organism studied by many scientists so that researchers can more easily compare their results and begin to unravel the properties of a given species.

moderately repetitive sequences sequences that are found a few hundred to several thousand times in the genome.

molecular clock the phenomenon by which the rate of neutral mutations can be used as a tool to measure evolutionary time.

molecular evolution the molecular changes in the genetic material that underlie the changes associated with evolution.

molecular genetics an examination of DNA structure and function at the molecular level.

molecular level with regard to gene expression, a level of observation in which genes affect the molecular properties of an organism.

molecular marker a segment of DNA that is found at a specific site in the genome and has properties that enable it to be uniquely recognized using molecular tools such as gel electrophoresis.

molecular pharming a recombinant technology that involves the production of medically important proteins in the mammary glands of livestock.

molecular profiling methods that enable researchers to understand the molecular changes that occur in diseases such as cancer.

molecular trait a trait that is observed at the molecular level. An example is the amount of a given protein in a cell.

monoallelic expression in the case of imprinting, refers to the phenomenon that only one of the two alleles of a given gene is transcriptionally expressed.

monohybrid a heterozygous individual produced from a single-factor cross.

monomorphic a term used to describe a gene that is found as only one allele in a population.

monophyletic group see *clade.*

monoploid an organism with a single set of chromosomes within its somatic cells.

monosomic a diploid cell that is missing a chromosome (i.e., $2n - 1$).

monozygotic twins twins that are genetically identical because they were formed from the same sperm and egg.

morph a form or phenotype in a population. For example, red eyes and white eyes are different eye color morphs.

morphogen a molecule that conveys positional information and promotes developmental changes.

morphological trait a trait that affects the morphology (physical form) of an organism. An example is eye color.

mosaicism when the cells of part of an organism differ genetically from the rest of the organism.

motif the name given to a domain or amino acid sequence that functions in a similar manner in many different proteins.

mouse model refers to a strain of mice that has been genetically altered to mimic a human genetic disease.

M phase a general name given to nuclear division that can apply to mitosis or meiosis. It is divided into prophase, prometaphase, metaphase, anaphase, and telophase.

mRNA see *messenger RNA.*

mtDNA an abbreviation for mitochondrial DNA.

MTOC see *microtubule-organizing center.*

multinomial expansion equation an equation to solve genetic problems involving three or more types of unordered events.

multiple alleles when the same gene exists in two or more alleles within a population.

multiplication method a method for solving independent assortment problems in which the probabilities of the outcome for each gene are multiplied together.

multipotent a type of stem cell that can differentiate into several different types of cells.

mutable site a site in a chromosome that tends to break at a fairly high rate due to the presence of a transposable element.

mutagen an agent that causes alterations in the structure of DNA.

mutant alleles alleles that have been created by altering a wild-type allele by mutation.

mutation a permanent change in the genetic material that can be passed from cell to cell or from parent to offspring.

mutation frequency the number of mutant genes divided by the total number of genes within the population.

mutation rate the likelihood that a gene will be altered by a new mutation.

myogenic bHLH protein a type of transcription factor involved in muscle cell differentiation.

N

n an abbreviation that designates the number of chromosomes in a set. In humans, $n = 23$, and a diploid cell has $2n = 46$ chromosomes.

narrow sense heritability heritability that takes into account only those genetic factors that are additive.

natural selection refers to the process whereby differential fitness acts on the gene pool. When a mutation creates a new allele that is beneficial, the allele may become prevalent within future generations because the individuals possessing this allele are more likely to reproduce and pass the beneficial allele to their offspring.

natural transformation a natural process of transformation that occurs in certain strains of bacteria.

negative control transcriptional regulation by repressor proteins.

negative frequency-dependent selection a pattern of natural selection in which the fitness of a genotype decreases when its frequency becomes higher.

neutral mutation a mutation that has no detectable effect on protein function or no detectable effect on the survival of the organism.

neutral theory of evolution the theory that most genetic variation observed in natural populations is due to the accumulation of neutral mutations.

next-generation sequencing technologies newer DNA sequencing technologies that are more rapid and inexpensive.

nick translation the phenomenon in which DNA polymerase uses its 5′ to 3′ exonuclease activity to remove a region of DNA and, at the same time, replaces it with new DNA.

nitrogen mustard an alkylating agent that can cause mutations in DNA.

nitrous acid a type of chemical mutagen that deaminates bases, thereby changing amino groups to keto groups.

nonallelic homologous recombination recombination that occurs at nonhomologous sites within chromosomes due to the occurrence of repetitive sequences.

nonautonomous element a transposable element that lacks a gene such as transposase or reverse transcriptase that is necessary for transposition.

noncoding strand the strand of DNA within a structural gene that is complementary to the mRNA. The noncoding strand is used as a template to make mRNA.

noncomplementation the phenomenon in which two mutant alleles in the same organism do not produce a wild-type phenotype.

non-Darwinian evolution see *neutral theory of evolution.*

nondisjunction event in which chromosomes do not segregate properly during mitosis or meiosis.

nonessential genes genes that are not absolutely required for survival, although they are likely to be beneficial to the organism.

nonhomologus end joining the joining of the broken ends of chromosomes via nonhomologous end-joining proteins.

nonhomologous end-joining (NHEJ) protein a protein that joins the ends of DNA fragments that are not homologous. This occurs during site-specific recombination of immunoglobulin genes.

nonhomologous recombination the exchange of DNA between nonhomologous segments of chromosomes or plasmids.

non-LTR retrotransposon a type of retrotransposon that does not have long terminal repeats.

nonneutral mutation a mutation that affects the phenotype of the organism and can be acted on by natural selection.

nonparental see *recombinant.*

nonrecombinant in a testcross, refers to a phenotype or arrangement of alleles on a chromosome that is not found in the parental generation.

nonsense codon a stop codon.

nonsense mutation a mutation that involves a change from a sense codon to a stop codon.

nontemplate strand a strand of DNA that is not used as a template during transcription.

nonviral-like retroelement a type of retroelement in which the sequence does not resemble a modern virus.

normal distribution a distribution for a large sample in which the trait of interest varies in a symmetrical way around an average value.

norm of reaction the effects of environmental variation on an individual's traits.

Northern blotting a technique used to detect a specific RNA within a mixture of many RNA molecules.

N-terminus see *amino terminus.*

nuclear genes genes that are located on chromosomes found in the cell nucleus of eukaryotic cells.

nuclear lamina a collection of fibers that line the inner nuclear membrane.

nuclear matrix a group of proteins that anchor the loops found in eukaryotic chromosomes.

nucleic acid RNA or DNA. A macromolecule that is composed of repeating nucleotide units.

nucleoid a darkly staining region that contains the genetic material of mitochondria, chloroplasts, or bacteria.

nucleolus a region within the nucleus of eukaryotic cells where the assembly of ribosomal subunits occurs.

nucleoprotein a complex of DNA (or RNA) and protein.

nucleoside structure in which a base is attached to a sugar, but no phosphate is attached to the sugar.

nucleosome the repeating structural unit within eukaryotic chromatin. It is composed of double-stranded DNA wrapped around an octamer of histone proteins.

nucleosome-free region (NFR) a region within a chromosome where nucleosomes are not found.

nucleotide the repeating structural unit of nucleic acids, composed of a sugar, phosphate, and base.

nucleotide excision repair (NER) a DNA repair system in which several nucleotides in the damaged strand are removed from the DNA and the undamaged strand is used as a template to resynthesize a normal strand.

nucleus a membrane-bound organelle in eukaryotic cells where the linear sets of chromosomes are found.

null hypothesis a hypothesis that assumes there is no real difference between the observed and expected values.

O

Okazaki fragments short segments of DNA that are synthesized in the lagging strand during DNA replication.

oligonucleotide a short strand of DNA, typically a few or a few dozen nucleotides in length.

oncogene a gene that promotes cancer due to a gain-of-function mutation.

one-gene/one-enzyme hypothesis the idea, which later needed to be expanded, that one gene encodes one enzyme.

oogenesis the production of egg cells.

open complex the region of separation of two DNA strands produced by RNA polymerase during transcription.

open conformation a loosely packed chromatin structure that is capable of transcription.

open reading frame (ORF) a genetic sequence that does not contain stop codons.

operator (or **operator site**) a sequence of nucleotides in bacterial DNA that provides a binding site for a genetic regulatory protein.

operon an arrangement in DNA in which two or more structural genes are found within a regulatory unit that is under the transcriptional control of a single promoter.

ORF see *open reading frame.*

organelle a large specialized structure within a cell, which is surrounded by a single or double membrane.

organism level when the level of observation or experimentation involves a whole organism.

organizing center in plants, a region of the meristem that ensures the proper organization of the meristem and preserves the correct number of actively dividing stem cells.

orientation-independent refers to certain types of genetic regulatory elements that can function in the forward or reverse direction. Certain enhancers are orientation independent.

origin of replication a nucleotide sequence that functions as an initiation site for the assembly of several proteins required for DNA replication.

origin of transfer the location on an F factor or within the chromosome of an Hfr strain that is the initiation site for the transfer of DNA from one bacterium to another during conjugation.

origin recognition complex (ORC) a complex of six proteins found in eukaryotes that is necessary to initiate DNA replication.

ortholog homologous genes in different species that were derived from the same ancestral gene.

outbreeding mating between genetically unrelated individuals.

outgroup in cladistics, a species or group of species that is distinctly different from the ingroup.

ovary (1) in plants, the structure in which the ovules develop; (2) in animals, the structure that produces egg cells and female hormones.

overdominance an inheritance pattern in which a heterozygote is more vigorous than either of the corresponding homozygotes.

ovule the structure in higher plants where the female gametophyte (i.e., embryo sac) is produced.

ovum a female gamete, also known as an egg cell.

oxidative DNA damage changes in DNA structure that are caused by reactive oxygen species (ROS).

oxidative stress an imbalance between the production of reactive oxygen species (ROS) and an organism's ability to break them down.

P

p an abbreviation for the short arm of a chromosome.

pachytene the third stage of prophase of meiosis I.

pair-rule gene one category of segmentation genes.

palindromic when a sequence is the same in the forward and reverse direction.

pangenesis an incorrect hypothesis of heredity. It suggested that hereditary traits could be modified depending on the lifestyle of the individual. For example, it was believed that a person who practiced a particular skill would produce offspring that would be better at that skill.

paracentric inversion an inversion in which the centromere is found outside of the inverted region.

paralogs homologous genes within a single species that constitute a gene family.

parapatric speciation (Greek, *para*, "beside"; Latin, *patria*, "homeland") a form of speciation that occurs when members of a species are only partially separated or when a species is very sedentary.

parasegments transient subdivisions that occur in the *Drosophila* embryo prior to the formation of segments.

parental in a testcross, refers to a phenotype or arrangement of alleles on a chromosome that is the same as one or both members of the parental generation.

parental generation (P generation) in a genetic cross, the first generation in the experiment. In Mendel's studies, the parental generation was true-breeding with regard to particular traits.

parental strand in DNA replication, the DNA strand that is used as a template.

parthenogenesis the formation of an individual from an unfertilized egg.

particulate theory of inheritance a theory proposed by Mendel. It states that traits are inherited as discrete units that remain unchanged as they are passed from parent to offspring.

P1 artificial chromosome (PAC) an artificial chromosome developed from P1 bacteriophage chromosomes.

paternal leakage the phenomenon in which species where maternal inheritance is generally observed, the male parent may, on rare occasions, provide mitochondria or chloroplasts to the zygote.

pattern the spatial arrangement of different regions of the body. At the cellular level, the body pattern is due to the arrangement of cells and their specialization.

pattern recognition in bioinformatics, this term refers to a program that recognizes a pattern of symbols.

PCR see *polymerase chain reaction.*

pedigree analysis a genetic analysis using information contained within family trees. In this approach, the aim is to determine the type of inheritance pattern that a gene follows.

pellet a collection of particles found at the bottom of a centrifuge tube.

peptide bond a covalent bond formed between the carboxyl group in one amino acid in a polypeptide chain and the amino group in the next amino acid.

peptidyl site (P site) a site on the ribosome that carries a tRNA along with a polypeptide chain.

peptidyl transfer the step during the elongation stage of translation in which the polypeptide is removed from the tRNA in the P site and transferred to the amino acid at the A site.

peptidyltransferase a complex that functions during translation to catalyze the formation of a peptide bond between the amino acid in the A site of the ribosome and the growing polypeptide chain.

pericentric inversion an inversion in which the centromere is located within the inverted region of the chromosome.

peripheral zone in plants, an area in the meristem that contains dividing cells that eventually differentiate into plant structures.

personalized medicine the application of genetic or molecular data in the treatment of disease.

petites mutant strains of yeast that form small colonies due to defects in mitochondrial function.

PGD see *preimplantation genetic diagnosis.*

P generation the parental generation in a genetic cross.

phage see *bacteriophage.*

phage λ a bacteriophage that infects *E. coli*.

pharming see *molecular pharming.*

phenetic approach an approach to construct a phylogenetic tree by considering the overall similarities among a group of the species without trying to understand their evolutionary history.

phenotype the observable traits of an organism.

phenylketonuria (PKU) a human genetic disorder arising from a defect in phenylalanine hydroxylase.

phosphodiester linkage in a DNA or RNA strand, a linkage in which a phosphate group connects two sugar molecules together.

photolyase an enzyme found in yeast and plants that can repair thymine dimers by splitting the dimers, which returns the DNA to its original condition.

photoreactivation a type of DNA repair of thymine dimers via photolyase that requires light.

phylogenetic tree a diagram that describes the evolutionary relationships among different species.

phylogeny the sequence of events involved in the evolutionary development of a species or group of species.

physical mapping the mapping of genes or other genetic sequences using DNA cloning methods.

physiological trait a trait that affects a cellular or body function. An example is the rate of glucose metabolism.

PKU see *phenylketonuria.*

plaque a clear zone within a bacterial lawn on a petri plate. It is due to repeated cycles of viral infection and bacterial lysis.

plasmid a general name used to describe circular pieces of DNA that exist independently of the chromosomal DNA. Some plasmids are used as vectors in cloning experiments.

pleiotrophy the multiple effects of a single gene on the phenotype of an organism.

pluripotent a type of stem cell that can differentiate into all or nearly all the types of cells of the adult organism.

point mutation a change in a single base pair within DNA.

polarity in genetics, the phenomenon in which a nonsense mutation in one gene affects the translation of a downstream gene in an operon.

polar mutation a mutation in an operon that affects the expression of another gene that is found downstream in the operon.

pollen also, pollen grain; the male gametophyte of flowering plants.

polyA-binding protein a protein that binds to the 3′ polyA tail of mRNAs and protects the mRNA from degradation.

polyadenylation the process of attaching a string of adenine nucleotides to the 3′ end of eukaryotic mRNAs.

polyA tail the string of adenine nucleotides at the 3′ end of eukaryotic mRNAs.

polycistronic mRNA an mRNA transcribed from an operon that encodes two or more proteins.

polygenic inheritance the transmission of traits that are governed by two or more different genes.

polymerase chain reaction (PCR) the method for amplifying a DNA region involving the sequential use of oligonucleotide primers and *Taq* polymerase.

polymerase switch during DNA replication, when one type of DNA polymerase (e.g., α) is switched for another type (e.g., β).

polymorphic a term used to describe a trait or gene that is found in two or more forms in a population.

polymorphism (1) the prevalence of two or more phenotypic forms in a population; (2) the phenomenon in which a gene exists in two or more alleles within a population.

polynucleotide a linkage of several nucleotides together to form a strand.

polypeptide a sequence of amino acids that is the product of mRNA translation. One or more polypeptides fold and associate with each other to form a functional protein.

polyploid an organism or cell with three or more sets of chromosomes.

polyribosome an mRNA transcript that has many bound ribosomes in the act of translation.

polysome see *polyribosome.*

polytene chromosome chromosomes that are found in certain cells, such as *Drosophila* salivary cells, in which the chromosomes have replicated many times and the copies lie side by side.

population a group of individuals of the same species that are capable of interbreeding with one another.

population genetics the field of genetics that is primarily concerned with the prevalence of genetic variation within populations.

population level when the level of observation or experimentation involves a population of organisms.

positional cloning a cloning strategy in which a gene is cloned based on its mapped position along a chromosome.

positional information chemical substances and other environmental cues that enable a cell to deduce its position relative to other cells.

position effect a change in phenotype that occurs when the position of a gene is changed from one chromosomal site to a different location.

positive control genetic regulation by activator proteins.

positive interference the phenomenon in which a crossover that occurs in one region of a chromosome decreases the probability that another crossover will occur nearby.

posttranslational describes events that occur after translation is completed.

posttranslational covalent modification the covalent attachment of a molecule to a protein after it has been synthesized via ribosomes.

postzygotic isolating mechanism a mechanism of reproductive isolation that prevents an offspring from being viable or fertile.

preimplantation genetic diagnosis (PGD) a form of genetic testing in which an embryo obtained via in vitro fertilization is tested for genetic abnormalities.

preinitiation complex the stage of transcription in which the assembly of RNA polymerase and general transcription factors occurs at the core promoter, but the DNA has not yet started to unwind.

pre-mRNA in eukaryotes, the transcription of structural genes produces a long transcript known as pre-mRNA, which is located within the nucleus. This pre-mRNA is usually altered by splicing and other modifications before it exits the nucleus.

prereplication complex (preRC) in eukaryotes, an assembly of at least 14 different proteins, including a group of 6 proteins called the origin recognition complex (ORC), that acts as the initiator of eukaryotic DNA replication.

prezygotic isolating mechanism a mechanism for reproductive isolation that prevents the formation of a zygote.

Pribnow box the TATAAT sequence that is often found at the -10 region of bacterial promoters.

primary structure with regard to proteins, the linear sequence of amino acids.

primase an enzyme that synthesizes a short RNA primer during DNA replication.

primer a short segment of DNA or RNA that initiates DNA replication.

primer extension in PCR, the step during which DNA is made by extending the length of a DNA primer.

primitive character a trait that is shared with a distant ancestor.

primosome a multiprotein complex composed of DNA helicase, primase, and several accessory proteins.

principle of parsimony the principle that the preferred hypothesis is the one that is the simplest.

prion an infectious particle that causes several types of neurodegenerative diseases affecting humans and livestock. It is composed entirely of protein.

probability the chance that an event will occur in the future.

probe a molecule, such as an oligonucleotide, that is used to identify another molecule or a site within a chromosome.

processive enzyme an enzyme, such as RNA and DNA polymerase, which glides along the DNA and does not dissociate from the template strand as it catalyzes the covalent attachment of nucleotides.

product rule the probability that two or more independent events will occur is equal to the products of their individual probabilities.

proflavin a type of chemical mutagen that causes frameshift mutations.

prokaryotes (Greek, "prenucleus") another name for bacteria and archaea. The term refers to the fact that their chromosomes are not contained within a separate nucleus of the cell.

prometaphase the second phase of M phase. During this phase, the nuclear membrane vesiculates, and the mitotic spindle is completely formed.

promoter a sequence within a gene that initiates (i.e., promotes) transcription.

promoter bashing the approach of making deletions in the vicinity of a promoter as a way to identify the core promoter and regulatory elements.

proofreading the ability of DNA polymerase to remove mismatched bases from a newly made strand.

prophage phage DNA that has been integrated into the bacterial chromosome.

prophase the first stage of M phase. The chromosomes have already replicated and begin to condense. The mitotic spindle starts to form.

protandrous hemaphrodite a species in which an individual may first become a male and then transforms into a female.

protease an enzyme that digests the polypeptide backbone found in proteins.

protein a functional unit composed of one or more polypeptides.

protein microarray a small silica, glass, or plastic slide that is dotted with many different proteins.

proteome the collection of all proteins that a given cell or species can make.

proteomics the study of protein function at the genome level. It involves the study of many proteins simultaneously.

proto-oncogene a normal cellular gene that does not cause cancer but which may incur a gain-of-function mutation or become incorporated into a viral genome and thereby lead to cancer.

protoplast a plant cell without a cell wall.

prototroph a strain that does not need a particular nutrient supplemented in its growth medium or diet.

provirus viral DNA that has been incorporated into the chromosome of a host cell.

proximodistal axis in animals, an axis involving limbs in which the part of the limb attached to the trunk is located proximal, whereas the end of the limb is located distal.

pseudoautosomal genes genes that are located in the regions that are found on both the X and Y chromosome.

pseudoautosomal inheritance the inheritance pattern of genes that are found on both the X and Y chromosomes. Even though such genes are located physically on the sex chromosomes, their pattern of inheritance is identical to that of autosomal genes.

pseudodominance a pattern of inheritance that occurs when a single copy of a recessive allele is phenotypically expressed because the second copy of the gene has been deleted from the homologous chromosome.

P site see *peptidyl site*.

pulse-chase experiment an experimental strategy in which cells are given a radiolabeled compound for a brief period of time—a pulse—followed by an excess amount of unlabeled compound—a chase.

punctuated equilibrium an evolutionary theory proposing that species exist relatively unchanged for many generations. These long periods of equilibrium are punctuated by relatively short periods during which evolution occurs at a relatively rapid rate.

Punnett square a diagrammatic method in which the gametes that two parents can produce are aligned next to a square grid as a way to predict the types of offspring the parents will produce and in what proportions.

purine a type of nitrogenous base that has a double-ring structure. Examples are adenine and guanine.

***P* value** in a chi square table, the probability that the deviations between observed and expected values are due to random chance.

pyrimidine a type of nitrogenous base that has a single-ring structure. Examples are cytosine, thymine, and uracil.

pyrosequencing a type of next-generation DNA sequencing.

pyrrolysine an amino acid that may be incorporated into polypeptides during translation.

Q

q an abbreviation for the long arm of a chromosome.

QTL see *quantitative trait loci.*

QTL mapping the determination of the location of QTLs using mapping methods such as genetic crosses coupled with the analysis of molecular markers.

quantitative genetics the area of genetics concerned with traits that can be described in a quantitative way.

quantitative trait a trait, usually polygenic in nature, that can be described with numbers.

quantitative trait loci (QTLs) the locations on chromosomes where the genes that influence quantitative traits reside.

quaternary structure in proteins, refers to the binding of two or more polypeptides to each other to form a protein.

R

radial loop domains organization of chromatin during interphase into loops, often 25,000 to 200,000 bp in size, which are anchored to the nuclear matrix at matrix-attachment regions.

RAG1/RAG2 enzymes that recognize recombination signal sequences and make double-stranded cuts. In the case of V/J recombination in immunoglobulin genes, a cut is made at the end of one V region and the beginning of one J region.

random genetic drift see *genetic drift.*

random mutation theory according to this theory, mutations are a random process—they can occur in any gene and do not involve exposure of an organism to a particular condition that selects for specific types of mutations.

random sampling error the deviation between the observed and expected outcomes due to chance.

reactive oxygen species (ROS) products of oxygen metabolism in all aerobic organisms that can damage cellular molecules, including DNA, proteins, and lipids.

reading frame a series of codons beginning with the start codon as a frame of reference.

realized heritability a form of narrow sense heritability that is observed when selective breeding is practiced.

real-time PCR a method of PCR in which the synthesis of DNA is monitored in real time. It can quantitate the starting of amount of DNA in a sample.

recessive a trait or gene that is masked by the presence of a dominant trait or gene.

recessive epistasis a form of epistasis in which an individual must be homozygous for either recessive allele to mask a particular phenotype.

reciprocal crosses a pair of crosses in which the traits of the two parents differ with regard to sex. For example, one cross could be a red-eyed female fly and a white-eyed male fly, and the reciprocal cross would be a red-eyed male fly and a white-eyed female fly.

reciprocal translocation when two different chromosomes exchange pieces.

recombinant (1) refers to combinations of alleles or traits that are not found in the parental generation; (2) describes DNA molecules that are produced by molecular techniques in which segments of DNA are joined to each other in ways that differ from their original arrangement in their native chromosomal sites. The cloning of DNA into vectors is an example.

recombinant DNA molecules molecules that are produced in a test tube by covalently linking DNA fragments from two different sources.

recombinant DNA technology the use of in vitro molecular techniques to isolate and manipulate different pieces of DNA.

recombinant vector a vector that contains an inserted fragment of DNA, such as a gene from a chromosome.

recombination see *genetic recombination.*

recombinational repair DNA repair via homologous recombination.

recombination signal sequence a specific DNA sequence that is involved in site-specific recombination. Such sequences are found in immunoglobulin genes.

redundancy see *gene redundancy*.

regression analysis the approach of using correlated data to predict the outcome of one variable when given the value of a second variable.

regression coefficient in a regression analysis, a coefficient that describes the slope of the regression line.

regulatory sequence, or **element** a sequence of DNA (or possibly RNA) that binds a regulatory protein and thereby influences gene expression. Bacterial operator sites and eukaryotic enhancers and silencers are examples.

regulatory transcription factor a protein or protein complex that binds to a regulatory element and influences the rate of transcription via RNA polymerase.

relaxosome a protein complex that recognizes the origin of transfer in F factors and other conjugative plasmids, cuts one DNA strand, and aids in the transfer of the T DNA.

release factor a protein that recognizes a stop codon and promotes translational termination and the release of the completed polypeptide.

repetitive sequences DNA sequences that are present in many copies in the genome.

replica plating a technique in which a replica of bacterial colonies is transferred to a new petri plate.

replication see *DNA replication.*

replication fork the region in which two DNA strands have separated and new strands are being synthesized.

replicative transposition a form of transposition in which the end result is that the transposable element remains at its original site and also is found at a new site.

replicative transposon a transposable element that increases in number during the transposition process.

replisome a complex that contains a primosome and dimeric DNA polymerase.

repressible gene a gene that is regulated by a corepressor or inhibitor, which are small effector molecules that cause transcription to decrease.

repressor a regulatory protein that binds to DNA and inhibits transcription.

reproduction the production of daughter cells from a mother cell or the production of offspring from parents.

reproductive cloning the cloning of a eukaryotic organism, which may involve the use of genetic material from somatic cells.

reproductive isolation the inability of a species to successfully interbreed with other species.

repulsion in linkage studies, refers to the phenomenon that some of the dominant alleles are on one chromosome and some of them are on the homologous chromosome.

resolution the last stage of homologous recombination, in which the entangled DNA strands become resolved into two separate structures.

resolvase an endonuclease that makes the final cuts in DNA during the resolution phase of recombination.

response element see *regulatory sequence or element.*

restriction endonuclease (or **restriction enzyme**) an endonuclease that cleaves DNA. The restriction enzymes used in cloning experiments bind to specific base sequences and then cleave the DNA backbone at two defined locations, one in each strand.

restriction fragment length polymorphism (RFLP) genetic variation within a population in the lengths of DNA fragments that are produced when chromosomes are digested with particular restriction enzymes.

restriction mapping a technique for determining the locations of restriction endonuclease sites within a segment of DNA.

restriction point a point in the G_1 phase of the cell cycle that causes a cell to progress to cell division.

retroelement a type of transposable element that moves via an RNA intermediate.

retroposon see *retroelement.*

retrospective testing a procedure in which a researcher formulates a hypothesis and then collects observations as a way to confirm or refute the hypothesis. This approach relies on the observation and testing of already existing materials, which themselves are the result of past events.

retrotransposition a form of transposition in which the element is transcribed into RNA. The RNA is then used as a template via reverse transcriptase to synthesize a DNA molecule that is integrated into a new region of the genome via integrase.

retrotransposon see *retroelement.*

reverse genetics an experimental strategy in which researchers first identify the wild-type gene using cloning methods. The next step is to make a mutant version of the wild-type gene, introduce it into an organism, and see how the mutant gene affects the phenotype of the organism.

reverse mutation see *reversion.*

reverse transcriptase an enzyme that uses an RNA template to make a complementary strand of DNA.

reverse transcriptase PCR (RT-PCR) a modification of PCR in which the first round of replication involves the use of RNA and reverse transcriptase to make a complementary strand of DNA.

reversion a mutation that returns a mutant allele back to the wild-type allele.

R factor a type of plasmid found commonly in bacteria that confers resistance to a toxic substance such as an antibiotic.

RFLP see *restriction fragment length polymorphism.*

RFLP map a genetic map composed of many RFLP markers.

RFLP mapping the mapping of a gene or other genetic sequence relative to the known locations of RFLPs within a genome.

R group the side chain of an amino acid.

rho (ρ) a protein that is involved in transcriptional termination for certain bacterial genes.

rho-dependent termination transcriptional termination that requires the function of the rho protein.

rho-independent termination transcription termination that does not require the rho protein. It is also known as intrinsic termination.

ribonucleic acid (RNA) a nucleic acid that is composed of ribonucleotides. In living cells, RNA is synthesized via the transcription of DNA.

ribose the sugar found in RNA.

ribosomal RNA (rRNA) RNA that is a component of the ribosome.

ribosome a large macromolecular structure that acts as the catalytic site for polypeptide synthesis. The ribosome allows the mRNA and tRNAs to be positioned correctly as the polypeptide is made.

ribosome-binding site a sequence in bacterial mRNA that is needed to bind to the ribosome and initiate translation.

riboswitch a form of genetic regulation in which an RNA can exist in two different secondary conformations.

ribozyme an RNA molecule with enzymatic activity.

right-left axis in bilateral animals, this axis determines the two sides of the body relative to the anteroposterior axis.

R loop experimentally, a DNA loop that is formed because RNA is displacing it from its complementary DNA strand.

RNA see *ribonucleic acid.*

RNA editing the process in which a change occurs in the nucleotide sequence of an RNA molecule that involves additions or deletions of particular bases, or a conversion of one type of base to a different type.

RNA-induced silencing complex (RISC) the complex that mediates RNA interference.

RNA interference the phenomenon that double-stranded RNA targets complementary RNAs within the cell for degradation.

RNA polymerase an enzyme that synthesizes a strand of RNA using a DNA strand as a template.

RNA primer a short strand of RNA, made by primase, that is used to elongate a strand of DNA during DNA replication.

RNase an enzyme that cuts the sugar-phosphate backbone in RNA.

RNaseP a bacterial enzyme that is an endonuclease and cuts precursor tRNA molecules. RNaseP is a ribozyme, which means that its catalytic ability is due to the action of RNA.

RNA splicing the process in which pieces of RNA are removed and the remaining pieces are covalently attached to each other.

Robertsonian translocation the structure produced when two telocentric chromosomes fuse at their short arms.

root meristem an actively dividing group of cells that gives rise to root structures.

rRNA see *ribosomal RNA.*

.

S

satellite DNA in a density centrifugation experiment, a peak of DNA that is separated from the majority of the chromosomal DNA. It is usually composed of highly repetitive sequences.

SBS see *sequencing by synthesis.*

scaffold a collection of proteins that holds the DNA in place and gives chromosomes their characteristic shapes.

scaffold-attachment region (SAR) a site in the chromosomal DNA that is anchored to the nuclear matrix or scaffold.

SCE see *sister chromatid exchange.*

science a way of knowing about our natural world. The science of genetics allows us to understand how the expression of genes produces the traits of an organism.

scientific method a basis for conducting science. It is a process that scientists typically follow so that they may reach verifiable conclusions about the world in which they live.

scintillation counting a technique that permits a researcher to count the number of radioactive emissions from a sample containing a population of radioisotopes.

search by content in bioinformatics, an approach to predict the location of a gene because the nucleotide content of a particular region differs significantly (due to codon bias) from a random distribution.

search by signal in bioinformatics, an approach that relies on known sequences such as promoters, start and stop codons, and splice sites to help predict whether or not a DNA sequence contains a gene.

secondary structure a regular repeating pattern of molecular structure, such as the DNA double helix or the α helix and β sheet found in proteins.

sedimentation coefficient a measure of centrifugation that is normally expressed in Svedberg units (S). A sedimentation coefficient has the units of seconds: $1 \text{ S} = 1 \times 10^{-13}$ seconds.

segmental duplication a small segment of a chromosome that has a tandem duplication.

segmentation gene in animals, a gene, whose encoded product is involved in the development of body segments.

segment polarity gene one category of segmentation genes.

segments anatomical subdivisions that occur during the development of species such as *Drosophila*.

segregate when two things are kept in separate locations. For example, homologous chromosomes segregate into different gametes.

selectable marker a gene that provides a selectable phenotype in a cloning experiment. Many selectable markers are genes that confer antibiotic resistance.

selection coefficient one minus the fitness value.

selectionists scientists who oppose the neutral theory of evolution.

selection limit the phenomenon in which several generations of artificial selection results in a plateau where artificial selection is no longer effective.

selective breeding programs and procedures designed to modify the phenotypes in economically important species of plants and animals.

selenocysteine an amino acid that may be incorporated into polypeptides during translation.

self-fertilization fertilization that involves the union of male and female gametes derived from the same parent.

selfish DNA hypothesis the idea that transposable elements exist because they possess characteristics that allow them to multiply within the host cell DNA and inhabit the host without offering any selective advantage.

self-splicing refers to RNA molecules that can remove their own introns without the aid of other proteins or RNA.

semiconservative model the correct model for DNA replication in which the newly made double-stranded DNA contains one parental strand and one daughter strand.

semiconservative replication the net result of DNA replication in which the DNA contains one original strand and one newly made strand.

semilethal alleles lethal alleles that kill some individuals but not all.

semisterility when an individual has a lowered fertility.

sense codon a codon that encodes a specific amino acid.

sense strand the strand of DNA within a structural gene that has the same sequence as mRNA except that T is found in the DNA instead of U.

sequence complexity the number of times a particular base sequence appears throughout the genome of a given species.

sequence element in genetics, a sequence with a specialized function.

sequence recognition in bioinformatics, this term refers to a program that recognizes particular sequence elements.

sequence-tagged site (STS) a short segment of DNA, usually between 100 and 400 bp long, the base sequence of which is found to be unique within the entire genome. Sequence-tagged sites are identified by PCR.

sequencing see *DNA sequencing*.

sequencing by synthesis (SBS) a next-generation form of DNA sequencing in which the synthesis of DNA is directly monitored to deduce the base sequence.

sequencing ladder a series of bands on a gel that can be followed in order (e.g., from the bottom of the gel to the top of the gel) to determine the base sequence of DNA.

sex chromosomes a pair of chromosomes (e.g., X and Y in mammals) that determines sex in a species.

sex determination the factors that govern whether an individual develops into a male or female organism.

sex-influenced inheritance an inheritance pattern in which an allele is dominant in one sex but recessive in the opposite sex. In humans, pattern baldness is an example of a sex-influenced trait.

sex-limited inheritance an inheritance pattern in which a trait is found in only one of the two sexes. An example would be beard development in men.

sex-limited traits traits that occur in only one of the two sexes.

sex linkage the phenomenon that certain genes are found on one of the two types of sex chromosomes but not both.

sex-linked gene a gene that is located on one of the sex chromosomes.

sex pilus (pl. **pili**) a structure on the surface of bacterial cells that acts as an attachment site to promote the binding of bacteria to each other. The sex pilus provides a passageway for the movement of DNA during conjugation.

sexual dimorphism species in which the males and females are morphologically distinct.

sexual reproduction the process whereby parents make gametes (e.g., sperm and egg) that fuse with each other in the process of fertilization to begin the life of a new organism.

sexual selection natural selection that acts to promote characteristics that give individuals a greater chance of reproducing.

shared derived character a trait shared by a group of organisms but not by a distant common ancestor.

Shine-Dalgarno sequence a sequence in bacterial mRNAs that functions as a ribosomal binding site.

shoot meristem an actively dividing group of cells that gives rise to shoot structures.

short-interfering RNAs (siRNAs) small RNA molecules that silence the expression of specific mRNAs via RNA interference.

short interspersed elements. See SINEs.

short tandem repeat sequences (STRs) short DNA sequences that are repeated many times in a row. Often found in centromeric and telomeric regions.

shotgun sequencing a genome sequencing strategy in which DNA fragments to be sequenced are randomly generated from larger DNA fragments.

shuttle vector a cloning vector that can propagate in two or more different species, such as *E. coli* and yeast.

side chain in an amino acid, the chemical structure that is attached to the carbon atom (i.e., the α carbon) that is located between the amino group and carboxyl group.

sigma (σ) factor a transcription factor that recognizes bacterial promoter sequences and facilitates the binding of RNA polymerase to the promoter.

silencer a DNA sequence that functions as a regulatory element. The binding of a regulatory transcription factor to the silencer decreases the level of transcription.

silent mutation a mutation that does not alter the amino acid sequence of the encoded polypeptide even though the nucleotide sequence has changed.

similarity with regard to DNA, refers to a comparison of DNA sequences that have regions where the bases match up. Similarity may be due to homology.

simple Mendelian inheritance an inheritance pattern involving a simple, dominant/recessive relationship that produces observed ratios in the offspring that readily obey Mendel's laws.

simple translocation when one piece of a chromosome becomes attached to a different chromosome.

simple transposition a cut-and-paste mechanism for transposition in which a transposable element is removed from one site and then inserted into another.

SINEs in mammals, short interspersed elements that are less than 500 bp in length.

single-factor cross a cross in which an experimenter is following the outcome of only a single trait.

single-nucleotide polymorphism (SNP) a genetic polymorphism within a population in which two alleles of the gene differ by a single nucleotide.

single-stranded binding protein a protein that binds to both of the single strands of DNA during DNA replication and prevents them from re-forming a double helix.

siRNA see *short interfering RNAs.*

sister chromatid exchange (SCE) the phenomenon in which crossing over occurs between sister chromatids, thereby exchanging identical genetic material.

sister chromatids pairs of replicated chromosomes that are attached to each other at the centromere. Sister chromatids are genetically identical.

site-directed mutagenesis a technique that enables scientists to change the sequence of cloned DNA segments.

site-specific recombination when two different DNA segments break and rejoin with each other at a specific site. This occurs during the integration of certain viruses into the host chromosome and during the rearrangement of immunoglobulin genes.

small nuclear riboprotein see *snRNP*.

SMC proteins proteins that use energy from ATP to catalyze changes in chromosome structure.

snRNP a complex containing small nuclear RNA and a set of proteins, which are components of the spliceosome.

somatic cell any cell of the body except for germ-line cells that give rise to gametes.

somatic hypermutation a phenomenon that occurs in certain cells of the immune system in which the genes that encode immunoglobulins are subject to a high rate of mutation due to errors in DNA replication.

somatic mutation a mutation in a somatic cell.

SOS response a response to extreme environmental stress in which bacteria replicate their DNA using DNA polymerases that are likely to make mistakes.

Southern blotting a technique used to detect the presence of a particular genetic sequence within a mixture of many chromosomal DNA fragments.

speciation the process by which new species are formed via evolution.

species a group of organisms that maintains a distinctive set of attributes in nature.

species concepts different approaches for distinguishing species.

spectrophotometer a device used by researchers to determine how much radiation at various wavelengths a sample can absorb.

spermatids immature sperm cells produced from spermatogenesis.

spermatogenesis the production of sperm cells.

sperm cell a male gamete. Sperm are small and usually travel relatively far distances to reach the female gamete.

spindle see *mitotic spindle apparatus*.

spindle pole during cell division in eukaryotes, one of two sites in the cell where microtubules originate.

spliceosome a multisubunit complex that functions in the splicing of eukaryotic pre-mRNA.

splicing see *RNA splicing*.

splicing factor a protein that regulates the process of RNA splicing.

spontaneous mutation a change in DNA structure that results from random abnormalities in biological processes.

spores haploid cells that are produced by certain species such as fungi (i.e., yeast and molds).

sporophyte the diploid generation of plants.

SR protein a type of splicing factor.

stabilizing selection natural selection that favors individuals with an intermediate phenotype.

stamen the structure found in the flower of higher plants that produces the male gametophyte (i.e., pollen).

standard deviation a statistic that is computed as the square root of the variance.

start codon a 3-base sequence in mRNA that initiates translation. It is usually 5′-AUG-3′ and encodes methionine.

stem cell a cell that has the capacity to divide and to differentiate into one or more specific cell types.

steroid receptor a category of transcription factors that respond to steroid hormones. An example is the glucocorticoid receptor.

stigma the structure in flowering plants on which the pollen land and the pollen tube starts to grows so that sperm cells can reach the egg cells.

stop codon a 3-base sequence in mRNA that signals the end of translation of a polypeptide. The three stop codons are 5′–UAA–3′, 5′–UAG–3′, and 5′–UGA–3′.

strain a variety that continues to produce the same characteristic after several generations.

strand in DNA or RNA, nucleotides covalently linked together to form a long, linear polymer.

stripe-specific enhancer in *Drosophila*, a regulatory region that controls the expression of a gene so that it occurs only in a particular parasegment during early embryonic development.

STRs see *short tandem repeat sequences*.

structural gene a gene that encodes the amino acid sequence within a particular polypeptide or protein.

STS see *sequence-tagged site*.

subcloning the procedure of making smaller DNA clones from a larger one.

submetacentric describes a chromosome in which the centromere is slightly off center.

subspecies a population within a species that has some distinct characteristics that differ from other members of the species.

subunit this term may have multiple meanings. In a protein, each subunit is a single polypeptide.

sum rule the probability that one of two or more mutually exclusive events will occur is equal to the sum of their individual probabilities.

supercoiling see *DNA supercoiling*.

supergroups a relatively recent way for evolutionary biologists to subdivide the eukaryotic domain.

supernatant following centrifugation, the fluid that is found above the pellet.

suppressor (or **suppressor mutation**) a mutation at a second site that suppresses the phenotypic effects of another mutation.

SWI/SNF family a group of related proteins that catalyze chromatin remodeling.

sympatric speciation (Greek, *sym*, "together"; Latin, *patria*, "homeland") a form of speciation that occurs when members of a species diverge while occupying the same habitat within the same range.

synapomorphy see *shared derived character*.

synapsis the event in which homologous chromosomes recognize each other and then align themselves along their entire lengths.

synaptonemal complex a complex of proteins that promote the interconnection between homologous chromosomes during meiosis.

synonymous codons two different codons that specify the same amino acid.

synteny a group of genes that are found in the same order on the chromosomes of different species.

systematists biologists who study the evolutionary relationships among different species.

T

T an abbreviation for thymine.

tandem array (or **tandem repeat**) a short nucleotide sequence that is repeated many times in a row.

tandem mass spectrometry the sequential use of two mass spectrometers. It can be used to determine the sequence of amino acids in a polypeptide.

***Taq* polymerase** a thermostable form of DNA polymerase used in PCR experiments.

target-site primed reverse transcription (TPRT) a mechanism of DNA synthesis that occurs during the movement of certain types of retrotransposons.

TATA box a sequence found within eukaryotic core promoters that determines the starting site for transcription. The TATA box is recognized by a TATA-binding protein, which is a component of TFIID.

tautomer the forms of certain small molecules, such as bases, which can spontaneously interconvert between chemically similar forms.

tautomeric shift a change in chemical structure such as an alternation between the keto and enol forms of the bases that are found in DNA.

T DNA a segment of DNA found within a Ti plasmid that is transferred from a bacterium to infected plant cells. The T DNA from the Ti plasmid becomes integrated into the chromosomal DNA of the plant cell by recombination.

T-DNA vector a vector that is found in *Agrobacterium tumifaciens* that is used to insert genes into plant cells.

TE see *transposable element*.

telocentric describes a chromosome with its centromere at one end.

telomerase the enzyme that recognizes telomeric sequences at the ends of eukaryotic chromosomes and synthesizes additional numbers of telomeric repeat sequences.

telomerase reverse transcriptase (TERT) the enzyme within telomerase that uses RNA as a template to make DNA.

telomeres specialized DNA sequences found at the ends of linear eukaryotic chromosomes.

telophase the fifth stage of M phase. The chromosomes have reached their respective poles and decondense.

temperate phage a bacteriophage that usually exists in the lysogenic cycle.

temperature-sensitive allele an allele in which the resulting phenotype depends on the environmental temperature.

temperature-sensitive (ts) lethal allele an allele that is lethal at a certain environmental temperature.

temperature-sensitive (ts) mutant a mutant that has a normal phenotype at a permissive temperature, but a defective phenotype, such as failure to grow, at the nonpermissive temperature.

template DNA a strand of DNA that is used to synthesize a complementary strand of DNA or RNA.

template strand see *template DNA*.

terminal deficiency see *terminal deletion*.

terminal deletion when a segment is lost from the end of a linear chromosome.

termination (1) in transcription, the release of the newly made RNA transcript and RNA polymerase from the DNA; (2) in translation, the release of the polypeptide and the last tRNA and the disassembly of the ribosomal subunits and mRNA.

termination codon see *stop codon*.

termination sequences (ter sequences) in *E. coli*, a pair of sequences in the chromosome that bind a protein known as the termination utilization substance (Tus), which stops the movement of the replication forks.

terminator a sequence within a gene that signals the end of transcription.

TERT see *telomerase reverse transcriptase*.

tertiary structure the three-dimensional structure of a macromolecule, such as the tertiary structure of a polypeptide.

testcross an experimental cross between a recessive individual and an individual whose genotype the experimenter wishes to determine.

tetrad the association among four sister chromatids during meiosis.

tetraploid having four sets of chromosomes (i.e., $4n$).

TFIID a type of general transcription factor in eukaryotes that is needed for RNA polymerase II function. It binds to the TATA box and recruits RNA polymerase II to the core promoter.

thermocycler a device that automates the timing of temperature changes in each cycle of a PCR experiment.

30-nm fiber the association of nucleosomes to form a more compact structure that is 30 nm in diameter.

three-factor cross a cross in which an experimenter follows the outcome of three different traits.

threshold traits traits that are inherited quantitatively due to the contribution of many genes, but the traits themselves are expressed qualitatively.

thymine a pyrimidine base found in DNA. It base-pairs with adenine in DNA.

thymine dimer a DNA lesion involving a covalent linkage between two adjacent thymine bases in a DNA strand.

Ti plasmid a tumor-inducing plasmid found in *Agrobacterium tumefaciens*. It is responsible for promoting tumor formation after a plant has been infected.

tissue-specific gene a gene that is highly regulated and is expressed in a particular cell type.

TNRE see trinucleotide repeat expansion.

topoisomerase an enzyme that alters the degree of supercoiling in DNA.

topoisomers DNA conformations that differ with regard to supercoiling.

totipotent a cell that possesses the genetic potential to produce an entire individual. A somatic plant cell or a fertilized egg is totipotent.

TPRT see *target-site primed reverse transcription.*

trait any characteristic that an organism displays. Morphological traits affect the appearance of an organism. Physiological traits affect the ability of an organism to function. A third category of traits are those that affect an organism's behavior (behavioral traits).

***trans*-acting factor** a regulatory protein that binds to a regulatory element in the DNA and exerts a *trans* effect.

***trans* configuration** refers to gene regulation in which the genetic factors do not need to be located in the same chromosome to exert their effects, usually because one of the factors is a diffusible protein.

transcription the process of synthesizing RNA from a DNA template.

transcriptional start site the site in a gene where transcription begins.

transcription-coupled DNA repair a form of DNA repair that is initiated when RNA polymerase cannot transcribe over a damaged region.

transcription factors a broad category of proteins that influence the ability of RNA polymerase to transcribe DNA into RNA.

transcription-repair coupling factor (TRCF) a protein that recognizes when RNA polymerase is stalled over a damaged region of DNA and recruits DNA repair enzymes to fix the damaged site.

transduction a form of genetic transfer between bacterial cells in which a bacteriophage transfers bacterial DNA from one bacterium to another.

***trans*-effect** an effect on gene expression that occurs even though two DNA segments are not physically adjacent to each other. *Trans*-effects are mediated through diffusible genetic regulatory proteins.

transfection (1) when a viral vector is introduced into a bacterial cell; (2) the introduction of any type of recombinant DNA into a eukaryotic cell.

transfer RNA (tRNA) a type of RNA used in translation that carries an amino acid. The anticodon in tRNA is complementary to a codon in the mRNA.

transformation (1) when a plasmid vector or segment of chromosomal DNA is introduced into a bacterial cell; (2) when a normal cell is converted into a malignant cell.

transgene a gene from one species that is introduced into another species.

transgenic organism an organism that has DNA from another organism incorporated into its genome via recombinant DNA techniques.

transition a point mutation involving a change of a pyrimidine to another pyrimidine (e.g., C to T) or a purine to another purine (e.g., A to G).

translation the synthesis of a polypeptide using the codon information within mRNA.

translational regulatory protein a protein that regulates translation.

translational repressor a protein that binds to mRNA and inhibits its ability to be translated.

translesion synthesis (TLS) the synthesis of DNA over a template strand that harbors some type of DNA damage. This occurs via lesion-replicating polymerases.

translocation (1) when one segment of a chromosome breaks off and becomes attached to a different chromosome; (2) when a ribosome moves from one codon in an mRNA to the next codon.

translocation cross the structure that is formed when the chromosomes of a reciprocal translocation attempt to synapse during meiosis. This structure contains two normal (nontranslocated chromosomes) and two translocated chromosomes. A total of eight chromatids are found within the cross.

transposable element (TE) a small genetic element that can move to multiple locations within the chromosomal DNA.

transposase the enzyme that catalyzes the movement of transposable elements.

transposition the phenomenon of transposon movement.

transposon see *transposable element.*

transversion a point mutation in which a purine is interchanged with a pyrimidine, or vice versa.

trinucleotide repeat expansion (TNRE) a type of mutation that involves an increase in the number of tandemly repeated trinucleotide sequences.

triplex DNA a double-stranded DNA that has a third strand wound around it to form a triple-stranded structure.

triploid an organism or cell that contains three sets of chromosomes.

trisomic a diploid cell with one extra chromosome (i.e., $2n + 1$).

tRNA see *transfer RNA.*

trp repressor a protein that binds to the operator site of the *trp* operon and inhibits transcription.

true-breeding line a strain of a particular species that continues to produce the same trait after several generations of self-fertilization (in plants) or inbreeding.

tumor-suppressor gene a gene that functions to inhibit cancerous growth.

two-dimensional gel electrophoresis a technique to separate proteins that involves isoelectric focusing in the first dimension and SDS-gel electrophoresis in the second dimension.

two-factor cross a cross in which an experimenter follows the outcome of two different traits.

U

U an abbreviation for uracil.

unbalanced translocation a translocation in which a cell has too much genetic material compared with a normal cell.

unipotent a type of stem cell that can differentiate into only a single type of cell.

universal in genetics, the phenomenon that nearly all organisms use the same genetic code with just a few exceptions.

up promoter mutation a mutation in a promoter that increases the rate of transcription.

up regulation genetic regulation that leads to an increase in gene expression.

uracil a pyrimidine base found in RNA.

UTR an abbreviation for the untranslated region of mRNA.

U-tube a U-shaped tube that has a filter at bottom of the U. The pore size of the filter allows the passage of small molecules (e.g., DNA molecules) from one side of the tube to the other, but restricts the passage of bacterial cells.

V

variance the sum of the squared deviations from the mean divided by the degrees of freedom.

variants individuals of the same species that exhibit different traits. An example is tall and dwarf pea plants.

V(D)J recombination site-specific recombination that occurs within immunoglobulin genes.

vector a small segment of DNA that is used as a carrier of another segment of DNA. Vectors are used in DNA cloning experiments.

vertical evolution the phenomenon that species evolve from preexisting species by the accumulation of gene mutations and by changes in chromosome structure and number. Vertical evolution involves genetic changes in a series of ancestors that form a lineage.

vertical gene transfer the transfer of genetic material from parents to offspring or from mother cell to daughter cell.

viral envelope a structure composed of a membrane and viral spike proteins that surrounds the capsid of certain eukaryotic viruses.

viral genome the genetic material of a virus.

viral-like retroelements retroelements that are evolutionarily related to known retroviruses. These transposable elements have retained the ability to move around the genome, though, in most cases, mature viral particles are not produced.

viral reproductive cycle the series of steps that occur after a virus has infected a cell; in bacteriophages, this may include a lysogenic and lytic cycle.

viral vector a vector used in gene cloning that is derived from a naturally occurring virus.

virulent phage a phage that follows the lytic cycle.

virus a small infectious particle that contains nucleic acid as its genetic material, surrounded by a capsid of proteins. Some viruses also have an envelope consisting of a membrane embedded with spike proteins.

VNTRs also called minisatellites; segments of DNA that are located in several places in a genome and have a variable number of tandem repeats. The pattern of VNTRs was originally used in DNA fingerprinting.

W

Western blotting a technique used to detect a specific protein among a mixture of proteins.

whole-genome shotgun sequencing a genome sequencing strategy that bypasses the mapping step. The whole genome is subjected to shotgun sequencing.

wild type a genotype or phenotype that is common in a natural population.

wild-type allele an allele that is fairly prevalent in a natural population, generally greater than 1% of the population. For polymorphic genes, there is more than one wild-type allele.

wobble base the first base (from the 5′ end) in an anticodon. This term suggests that the first base in the anticodon can wobble a bit to recognize the third base in the mRNA.

wobble rules rules that govern the binding specificity between the third base in a codon and the first base in an anticodon.

X

X chromosomal controlling element (Xce) a region adjacent to Xic that influences the choice of the active X chromosome during the process of X inactivation.

X inactivation a process in which mammals equalize the expression of X-linked genes by randomly turning off one X chromosome in the somatic cells of females.

X-inactivation center (Xic) a site on the X chromosome that appears to play a critical role in X inactivation.

X-linked genes (alleles) genes (or alleles of genes) that are physically located within the X chromosome.

X-linked inheritance an inheritance pattern in certain species that involves genes that are located only on the X chromosome.

X-linked recessive pattern an allele or trait in which the gene is found on the X chromosome and the allele is recessive relative to a corresponding dominant allele.

Y

YAC see *yeast artificial chromosome.*

yeast artificial chromosome (YAC) a cloning vector propagated in yeast that can reliably contain very large insert fragments of DNA.

Y-linked genes (alleles) genes (or alleles of genes) that are located only on the Y chromosome.

Z

Z DNA a left-handed DNA double helix that is found occasionally in the DNA of living cells.

zygote a cell formed from the union of a sperm and egg.

zygotene the second stage of prophase of meiosis I.

zygotic gene a gene that is expressed after fertilization.

This page intentionally left blank

CREDITS

Photographs

Front Matter

i: ©Willard Clay/Oxford Scientific/Getty Images; xiii: © Photo courtesy of the College of Veterinary Medicine, Texas A&M University/Corbis; xiii: © Photomicrographs by Dr. Conly L. Rieder, Wadsworth Center, Albany, New York 12201-0509; v: © Ken Eward/Photo Researchers; vi: From Patrick Cramer, David A. Bushnell, Roger D. Kornberg. “Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution.” *Science*, Vol. 292:5523, 1863-1876, June 8, 2001. Reprinted with permission of AAAS.; v: © Argus Fotoarchive/Peter Arnold; vii: © Royalty-Free/Corbis; xiii: Courtesy Rob Brooker; xiv: © Biophoto Associates/Photo Researchers; xiv: © CNRI /Science Photo Library/Photo Researchers

Chapter 1

Opener: © Photo courtesy of the College of Veterinary Medicine, Texas A&M University/Corbis; 1.2(left): © Roslin Institute/AP Images; 1.3a(right): © Advanced Cell Technology, Inc., Worcester, Massachusetts; 1.3b: Photo taken by Flaminia Catteruccia, Jason Benton and Andrea Crisanti, and assembled by www.luciariccidesign.com.; 1.4: © Biophoto Associates/Photo Researchers; 1.5: © CNRI /Science Photo Library/Photo Researchers; 1.8: © Wildlife GmbH/Alamy; 1.9a: © Joseph Sohm; ChromoSohm Inc./Corbis; 1.9b: ©Paul Edmondson/Photodisc Red/Getty Images; 1.10: © March of Dimes Birth Defects Foundation; 1.13b: © SciMAT/Photo Researchers; 1.13a: © George Musil/Visuals Unlimited; 1.13c: © Steve Hopkin/ardea.com; 1.13d: © Brad Mogen/Visuals Unlimited; 1.13e: © Mark Smith/Photo Researchers; 1.13f: © J-M. Labat/Photo Researchers; 1.13g: © Wally Eberhart/Visuals Unlimited.

Chapter 2

Opener: © Photomicrographs by Dr. Conly L. Rieder, Wadsworth Center, Albany, New York 12201-0509; 2.2b: © Burger/Photo Researchers; 2.2c: © Leonard Lessin/Peter Arnold; 2.6a(right): © Biophoto Associates/Photo Researchers; 2.6a(left): © Leonard Lessin/Peter Arnold; 2.8a-f: © Photomicrographs by Dr. Conly L. Rieder, Wadsworth Center, Albany, New York 12201-0509; 2.9a(1): © Dr. David M. Phillips/Visuals Unlimited; 2.9b: © Ed Reschke; 2.11: © Diter von Wettstein.

Chapter 3

Opener: © Chris Martin Bahr/Science Photo Library/Photo Researchers; 3.1: © SPL/Photo Researchers; 3.2b: © Nigel Cattlin/Photo Researchers.

Chapter 4

Opener: © William Leaman/Alamy; 4.2a: © Scott Camazine/Photo Researchers; 4.2b: © Photo by Rob Flynn/USDA; 4.3a: © Nature Picture Library/Alamy; 4.3b: © cbpix/Alamy; 4.4a: © Rolf Nussbaumer Photography/Alamy; 4.4b(1): © Arco Images GmbH/Alamy; 4.4b(2): © Arco Images GmbH/Alamy; 4.5b: © Tim Davis/Photo Researchers; 4.5a-b© Courtesy of I.Solovei, University of Munich (LMU).; 4.11a: © AP Images.

Chapter 5

Opener: © Robert Calentine/Visuals Unlimited; 5.1: © blickwinkel/Alamy; 5.3b: © Bob Shanley/The Palm Beach Post; 5.4a(1): © Peter Weimann/Animals Animals; 5.4a(2): © Tom Walker/Visuals Unlimited; 5.4b: © Sally Haugen/Virginia Schuett, www.pkunews.org; 5.4c: © PHOTOTAKE Inc./Alamy; 5.7a-b: © Stan Flegler/Visuals Unlimited; 5.10a: © National Parks Service, Adams National Historical Park; 5.12a-b: © Robert Maier/Animals, Animals; 5.13a: © Jane Burton/naturepl.com.

Chapter 6

Opener: © John Mendenhall, Institute for Cellular and Molecular Biology, University of Texas at Austin; 6.1: Gibbs, SP., Mak, R., Ng, R., & Slankis, T. “The Chloroplast Nucleoid In Ochromonas Danica. II. Evidence For An Increase In Plastid DNA During Greening.” *J Cell Sci*. 1974 Dec; 16(3):579-91. Fig. 1. By permission of the Company of Biologists Limited.; 6.6: Reproduced with permission from *The Journal of Cell Biology*, 1977, 72:687-694. Copyright 1977 The Rockefeller University Press. Reprinted with permission.; 6.9: © Courtesy of Dr. Argiris Efstratiadis.

Chapter 7

Opener: © Educational Images Ltd/Custom Medical Stock Photo.

Chapter 8

Opener: © BSIP/Phototake; 8.1a(right): © Carlos R Carvalho/Universidade Federal de Viçosa.; 8.1a(left): © Scott Camazine /Photo Researchers; 8.1a(center): © Michael Abbey/Photo Researchers; 8.1c: © CNRI/Phototake; 8.4a: © Biophoto Assocates/Science Source/Photo Researchers; 8.4b: © Jeff Noneley; 8.13b: © Paul Benke/University of Miami School of Medicine; 8.13c: © Will Hart/PhotoEdit; 8.18a-b: © A. B. Sheldon; 8.19b: © David M. Phillips/Visuals Unlimited; 8.20a: © James Steinberg/Photo Researchers; 8.20b(both): © Biophoto Associates/Science Source/Photo Researchers; 8.26: Robinson, T.J. & Harley, E.H. “Absence of geographic chromosomal variation in the roan and sable antelope and the cytogenetics of a naturally occurring hybrid.” *Cytogenet Cell Genet*. 1995; 71(4): 363-9. Permission granted by S. Karger AG, Basel. Reprinted with permission.

Chapter 9

Opener: © David Scharf/Peter Arnold; 9.4b: © Dr. L. Caro/Science Photo Library/Photo Researchers.

Chapter 10

Opener: © Nigel Cattlin / Photo Researchers; 10.1a: © Robley C. Williams/Biological Photo Service; 10.1b: © Courtesy of R.C. Valentine and H.G. Pereira. Reprinted from *Journal of Molecular Biology*, Vol. 13, Number 2, R.C. Valentine and H.G. Pereira, Antigens and Structure of the Adenovirus, Pages 71-83, 1965, with permission from Elsevier.; 10.1c: © Chris Bjornberg/Photo Researchers; 10.1d: © Omikron/Photo Researchers; 10.4: © CDC/C. S. Goldsmith and A. Balish; 10.5: © CDC/Peter Arnold; 10.8(left): © Carolina Biological Supply/Phototake; 10.8(right): © Carolina Biological Supply/Phototake.

Chapter 11

Opener: © Ken Eward /Photo Researchers; 11.3(left): © Omikron/Photo Researchers; 11.11: © Pictorial Parade/Getty Images; 11.12a: From *The Double Helix*. By James D. Watson, 1968, Atheneum Press, NY. By permissions of Cold Spring Harbor Laboratory Archives; 11.13a: © Barrington Brown/Photo Researchers; 11.13b: © Hulton|Archive by Getty Images; 11.15b: © Laguna Design/Photo Researchers; 11.19b: © Alfred Pasieka/Photo Researchers.

Chapter 12

Opener: © Dr. Gopal Murti/Visuals Unlimited; 12.2: © American Society for Microbiology.; 12.8: © Simpson’s Nature Photography; 12.8c: © William Leonard; 12.12a-b: This article was published in *Cell*. 12(1). Thoma , F. & Koller, T. “Influence of histone H1 on chromatin structure.” p. 103. f. 2a&c, Copyright Elsevier, 1977. Reproduced with permission.; 12.13a: © Jerome Rattner/University of Calgary; 12.14b-c: Nickerson et al. “The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus.” *PNAS*, 94: 4446-4450. Figure 3a &b. © Copyright 1997 National Academy of Sciences, U.S.A.; 12.17a: © Dr. Gopal Murit/Visuals Unlimited, 12.17b (Rt): © Jerome Rattner/University of Calgary; 12.17c: This article was published in *Cell*. Nov;12(3). Paulson, JR. & Laemmli, UK. “The structure of histone-depleted metaphase chromosomes.”; 817-28. f. 5. Copyright Elsevier, 1977.Reproduced with permission.; 12.17c-d : © Peter Engelhardt/Department of Virology, Haartman Institute.; 12.18a: © Peter Engelhardt/Department of Virology, Haartman Institute.; 12.18b: © Dr. Donald Fawcett /Visuals Unlimited.

Chapter 13

Opener: © Clive Freeman, The Royal Institution/Photo Researchers; 13.1: Meselson M, Stahl, F. "The Replication of DNA in Escherichia Coli." *PNAS*. Vol. 44, 1958. f. 4a, p. 673. Courtesy of M. Meselson.; 13.4b: From *Cold Spring Harbor Symposia of Quantitative Biology*, 28, p. 43 (1963). Copyright holder is Cold Spring Habour Laboratory Press.; 13.17: This article was published in *Journal of Molecular Biology*. Mar 14;32(2). Huberman JA. Riggs AD. "Links On the mechanism of DNA replication in mammalian chromosomes." 327-41. Copyright Elsevier, 1968. Reprinted with permission.; 13.24b: © Dr. John D. Cunningham/ Visuals Unlimited.

Chapter 14

Opener: From Patrick Cramer, David A. Bushnell, Roger D. Kornberg. "Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution." *Science*, Vol. 292:5523, 1863-1876, June 8, 2001. Reprinted with permission of AAAS.; 14.12a(left): From Seth Darst, Bacterial RNA polymerase. *Current Opinion in Structural Biology*. Reprinted with permission of the author.; 14.12a(right): From Patrick Cramer, David A. Bushnell, Roger D. Kornberg. "Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution." *Science*, Vol. 292:5523, 1863-1876, June 8, 2001. Figure 1. Image courtesy of David Bushnell.

Chapter 15

Opener: © Tom Pantages; 15.14a: © E. Kiseleva and Donald Fawcett/Visuals Unlimited; 15.14b: © Tom Pantages; 15.20: Miller , O. L. *Scientific American*. Vol 228:3, 1973. p.35. Image courtesy O.L. Miller.

Chapter 16

Opener: © Illustration by David S. Goodsell of The Scripps Research Institute and the RCSB PDB.; 16.10b: © Mitchell Lewis, University Of Pennsylvania Medical Center/Science Photo Library/Photo Researchers; 18.20: © Dr. Kenneth Greer/Visuals Unlimited.

Chapter 17

Opener: Zif268 zinc fingers bound to DNA (Pavletich and Pabo, 1991). Courtesy Song Tan, Penn State University, <http://www.bmb.psu.edu/faculty/tan/lab>.

Chapter 18

Opener: © Robert Brooker; 18.1a-b: © Phototake/Alamy; 18.3b: Aulner et al. "The AT-Hook Protein D1 Is Essential for Drosophila melanogaster Development and Is Implicated in Position-Effect Variegation." *Molecular and Cellular Biology*. 22: 4. 1218-1232, February 2002. Figure. 7. © 2002 American Society for Microbiology. Images courtesy of E. Kas.; 18.5: © Scott Aitken, www.scottpix.com.

Chapter 19

Opener: © Argus Fotoarchive/Peter Arnold.

Chapter 20

Opener: © Najlah Feanny/Saba/Corbis; 20.2: R. L. Brinster and R. E. Hammer, School of Veterinary Medicine, University of Pennsylvania.; 20.4: Glofish ©www.glofish.com; 20.5: © Alan Handyside, Wellcome Images; 20.9: © Photo courtesy of the College of Veterinary Medicine, Texas A&M University/Corbis; 20.13: © Jack Bostrack/Visuals Unlimited; 20.15: Courtesy Monsanto; 20.16: Bill Barksdale/AGStockUSA.

Chapter 21

Opener: © Dr. Peter Lansdorp/Visuals Unlimited; 21.3: From Ried, T., Baldini, A., Rand, T.C., and Ward, D.C. "Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. *PNAS*. 89: 4.1388-92. 1992. Courtesy Thomas Ried.

Chapter 22

Opener: © Alfred Pasieka/Photo Researchers; 22.4b: © Medical School. University of Newcastle upon Tyne/Simon Fraser/Photo Researchers.

Chapter 23

Opener: © Royalty-Free/Corbis; 23.1(top): © Dr. P. Marazzi/Photo Researchers; 23.1(middle): © Hiroya Minakuchi /Seapics.com; 23.1(bottom): © Mitch Reardon/Photo Researchers; 23.4: © Hulton-Deutsch Collection/Corbis; 23.8: © CNRI/Science Photo Library/Photo Researchers; 23.18a-b: © Courtesy Dr. Ruhong Li, Molecular and Cell Biology, University of California at Berkeley. Permission granted by The Duesberg Laboratory. Appeared in *Science*, Vol. 297, Number 5581, Issue 26, Jul 2002, p. 544. Author: Jean Marx, Title: "Debate Surges Over the Origins of Genomic Defects in Cancer.".

Chapter 24

Opener: © N. A. Callow/NHPA/Photoshot; 24.3: E. B. Lewis., California Institute of Technology Archives.; 24.8a-b-24.11(1): Christiane Nusslein-Volhard, *Development*, Supplement 1, 1991. © The Company of Biologists Limited; 24.11(2-3): Jim Langeland, Steve Paddock and Sean Carroll/University of Wisconsin - Madison; 24.13a: © Juergen Berger/Photo Researchers; 24.13b: © F. R. Turner, Indiana University/Visuals Unlimited; 24.17: © Photodisc/Object Series/Vol. 50/Getty Images RF; 24.20: © Jeremy Burgess/Photo Researchers; 24.23a-c: Elliott Meyerowitz and John Bowman. *Development*.Vol. 112:1-20, 1991. Courtesy of Elliott Meyerowitz.; p. 610: © Courtesy Stephen Small/New York University.

Chapter 25

Opener: © PhotoDisc/Getty Images RF; 25.1a: © Miguel Castro/Photo Researchers; 25.1b: © D.Parer & E.Parer-Cook/ardea.com; 25.2 (all): © Geoff Oxford; 25.11: © OSF/Animals Animals; 25.13: © Paul Harcourt Davies/SPL/ Photo Researchers; 25.14: © Gerald & Buff Corsi/Visuals Unlimited; 25.17: © Corbis RF; 25.21: © Leonard Lessin/Peter Arnold.

Chapter 26

Opener: © PhotoDisc/Vol. 74/Getty Images RF; ; 26.1: © From Albert and Blakeslee, Corn and Man, *Journal of Heredity*. 1914, Vol. 5, pg. 51. Oxford University Press. This image is in public domain.: 26.3(bottom): © Nigel Cattlin/Visuals Unlimited; 26.3(top): © Grant Heilman Photography/Alamy; 26.9(top left): © Henry Ausloos/Animals Animals; 26.9(top right): © Roger Tidman/Corbis; 26.9(bottom left): © Philip Gould/Corbis; 26.9(bottom right): © Bildarchiv/Alamy; 26.10(mustard): © Steven P. Lynch RF; 26.10(Kohlrabi): © Burke Triolo Productions/Getty Images RF; 26.10(kale): © D. Normark/ PhotoLink/Getty Images RF; 26.10(broccoli): © Comstock/PunchStock RF; 26.10(brussels sprouts): © Michael P. Gadomski/Photo Researchers; 26.10(cabbage): © Author's Image/PunchStock RF; 26.10(cauliflower): © Doug Wilson/USDA.

Chapter 27

Opener: © LWA/Stephen Welstead/Blend Images/Corbis RF; 27.1b(left): © Rod Planck/Photo Researchers; 27.1a(left): © Wildlife GmbH/Alamy; 27.1b(right): © Ron Austin/Photo Researchers; 27.3(left): © Paul & Joyce Berquist/Animals Animals; 27.3(right): © Gerald and Buff Corsi/Visuals Unlimited; 27.16a(both): © Courtesy Ed Laufer; 27.16b-c: Courtesy of Dr. J.M. Hurle. From *Development*. 1999 Dec;126(23):5515-22. Reproduced with permission.; 27.18a-b: © Prof. Walter J. Gehring, University of Basel.

Appendix

A1a: © Michael Gabridge/Visuals Unlimited; A1b: © Fred Hossler/Visuals Unlimited; A8: © Richard Wehr/Custom Medical Stock Photo.

Line Art

Chapter 1

1.1 a,b: Courtesy of the U.S. Department of Energy Genome Programs. http://genomics.energy.gov

Chapter 11

11.16: Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.

Chapter 14

14.12: Reprinted from "Bacterial RNA Polymerase" by Darst, S.A. Current Opinion in Structural Biology, vol. 11 (2) p. 157. Copyright 2001 with permission from Elsevier.

Chapter 16

16.10: From Crystal Structure of the Lactose Operon Repressor and its Complexes with DNA and Inducer by M. Lewis et al., Science, vol 271, pp.1247-1254, March 1, 1996. Reprinted with permission from AAAS.

INDEX

Page numbers in *italics* indicate figures or tables.

A

ABC model, 604, *605*
Abd-B gene, 596
ABO blood group, 96–98, *97*
Achondroplasia, *555*
Acquired immune deficiency syndrome (AIDS), 214. *See also* Human immunodeficiency virus (HIV)
Acridine dyes, 427, *427*
Acrocentric chromosome, 156, *157*
Activator protein, 356, *356,* 380, *381*
Acutely transforming viruses, 569–570
Adaptor hypothesis, 338, 341–342
Adenine, 5, 232, *232, 233,* 235, *235*
Adenosine, 232, *233*
Adenosine deaminase deficiency, 492, 493–495, *494, 554*
Adenosine diphosphate (ADP), 232, *233*
Adenosine monophosphate (AMP), 232, *233*
  cyclic (cAMP), 364–365, *365*
Adenosine triphosphate (ATP), 232, *233*
Adenovirus, *207, 208*
Adjacent-1 segregation, 165, 166, *166*
Adjacent-2 segregation, 166, *166*
Aedes aegypti, 623, *623*
Age, parental, 170–171, *171*
Agrobacterium radiobacter, 476, 488, *489, 490*
Agrobacterium tumefaciens, 476, 488, *489, 490*
Aicardi syndrome, *557*
Albinism, *90, 553, 554*
Alkaptonuria, 325, *325*
Alkyltransferase, 430–431, *431*
Allele(s), 7–8, 21. *See also* Gene(s)
  ABO, 96–98, *97*
  beneficial, 623, *623*
  definition of, 46
  directional selection of, 621–623, *622, 623*
  disease-causing, 558–563, *559, 560*
  dominant, 91
  incomplete dominance of, *94,* 94–95
  incomplete penetrance of, 91–92, *92*
  lethal, *89,* 100–101, *101*
  linked (*See* Linkage)
  loss-of-function, 14, 52, 103–105
  MN, 618–619
  multiple, 96
  mutant, 90–91, 414 (*See also* Mutation[s])
  recessive, *90,* 90–91
  semilethal, 100
  wild-type, 90, *90,* 414
  X-linked, 81
Allele frequency, 616–617
  directional selection and, *620, 621,* 621–623
  genetic drift and, 628–630, *629*
  in Hardy-Weinberg equation, 617–619, *618*
  migration and, 630–631
  mutation and, 634–635
Alligator, 73, *73*
Allis, C. David, 389
Allodiploidy, 176–179, *177, 178*
Allolactose, 358, *359,* 360, *361*
Allopatric speciation, 676–678, *677*
Alloploidy, 176, *177*
Allopolyploidy, 176, *177*
Allosteric enzyme, 372, *372*
Allosteric regulation, 357, 359
Allotetraploidy, 176–179, *177, 178*
Alopex lagopus, 92, *93*
Alper, Tikvah, 563
Alpha (α)–helix, 330–331, *333*
Alternate segregation, 165, 166, *166*
Alternative exon, 397, *397*
Alternative splicing, 316, *396,* 396–398, *397, 398, 536,* 536–537
Altschul, Stephen, 544
Alu sequence, 253
Ambros, Victor, 401–402
American holly, 74, *74*
Ames, Bruce, 429
Ames test, *429,* 429–430
Amino acids, 5. *See also* Polypeptide
  aliphatic, 329–330, *331*
  aromatic, *331*
  of lysozyme, 330, *332*
  nonstandard, 329, *331*
  polar, 330, *331*
  polypeptide incorporation of, 329–332, *331, 332, 333* (*See also* Translation)
  side chain (R group) of, 329–330
Aminoacyl (A) site, 343, *343*
Aminoacyl-tRNA synthetases, 340, *340*
2-Aminopurine, 427, *427*
Amino-terminal end (N-terminus), 329
Amish, 630
Ammospermophilus, 676–678, *677*
Amniocentesis, *562,* 562–563
amp^R gene, 445, 447–448, *448*
Anagenesis, 676, *677*
Anaphase
  of meiosis, *32,* 33
  of mitosis, *28,* 29
Ancestral (primitive) character, 681
Ancestral gene, 543–544, *544*
Androgenic insensitivity syndrome, *556*
Aneuploidy, *168, 169,* 169–171, *170, 171*
  cancer and, 574
  phenotype and, *170,* 170–171
Angelman syndrome, 120–121, *122*
Aniridia, *555*
Anopheles gambiae, 518
Antelope hybrid, 176–177, *178*
Antelope squirrels, 676–678, *677*
Antennapedia complex, 596, *596*
Anteroposterior axis, *590,* 591
Antibody, 605–607, *606, 607*
  as probe, 460–461, *461*
  transgenic production of, 482, *482*
Antibody microarray, 539
Anticipation, 424–425
Anticodon, 327
Antigen, 460, 605
Antisense RNA, 372, *372*
  in mRNA silencing, 400–401, *400–401*
α_1-Antitrypsin, transgenic production of, 482, *482*
Antp gene, 596, *596*
APC gene, 574, *575*
AP endonuclease, 431–432, *432*
Apolipoprotein B, 317–318
Apoptosis, 571, 586, *586*
Apple maggot fly, 678
Arabidopsis thaliana, 13, 380, *518,* 602–605, *603, 604*
Arctic fox, 92, *93*
ARS (autonomously replicating sequence) elements, 283
Artificial selection, *663,* 663–666, *664, 665*
Asexual reproduction, 23–24, *24*
Assortative mating, 631
Astrachan, Lazarus, 302–303
AT/GC rule, 235, *237,* 238
Attenuation, 367, 368–370, *369*
AUG codon, 345–346. *346*
Automated DNA fingerprinting, 637, *637*
Autonomous element, 520
Autopolyploidy, 176, *177*
Autoradiography, A-6
Autosomal dominant inheritance, 554–556, *555*
Autosomal recessive inheritance, 553–554, *554*
Autosomes, 72
Auxotroph, 187
Avery, Oswald, 227

B

Bacillus subtilis, 247, 373, *374*
Bacillus thuringiensis, 476, 489–490, *491*
Bacteria, 185–201. *See also Escherichia coli*
  auxotroph, 187
  biological control with, 476
  chromosomes of (*See* Bacterial chromosomes)
  DNA replication in (*See* DNA replication, bacterial)
  gene regulation in (*See* Gene regulation, bacterial)
  genetic transfer in, *186,* 186–187, 636, *636*
    conjugation, 186, *186, 187,* 187–190, *188, 189, 190*
    transduction, 186, *186,* 196–199, *197, 198*
    transformation, 186, *186,* 199–200, *200*
  genomes of, 517, *518,* 542
  Hfr strains of, *191,* 191–196, *192*
  insulin production by, 475, *476*
  merozygote, 362
  nucleoids of, 247, *247*
  plasmids in, 189–190
  pollutant reduction with, 476–477
  prototroph, 187
  reproduction of, 23–24, *24*
  riboswitch in, 373, *374*
  transcription in (*See* Transcription, bacterial)
  translation in (*See* Translation, bacterial)
Bacterial artificial chromosomes, 508–509, *509,* 512–513, *513*
Bacterial chromosomes, 246–247
  intergenic regions of, 246, *247*
  loop domains of, 248, *248*
  loops of, 247–248, *248*
  negative supercoiling of, 248–249, *249*
  origin of replication of, 247, *247*
  repetitive sequences of, 247, *247*
  structural genes of, 246, *247*
  structure of, *247,* 247–250, *248, 249*
  supercoiling of, *248,* 248–250, *249*
Bacterial conjugation, 186, *186,* 187–190
  for genetic mapping, 193–196, *194, 195*
  genetic transfer during, *187,* 187–188
    direct contact in, 188, *188*

- Bacterial conjugation, *continued*
    - F factor in, 188–189, *189, 190*
    - in *Hfr* strains, *191,* 191–196, *192*
    - plasmids in, 189–190
- Bacterial transduction, 196–199
    - bacteriophages in, 196–197, *197*
    - cotransduction in, 197–199, *198*
- Bacterial transformation, 199–200, *200*
    - artificial, 199
    - for genetic mapping, 200
    - natural, 199
    - in *Streptococcus pneumoniae, 226,* 226–228, *227*
- Bacteriophage, 196–197, *197,* 209. *See also* Phage
    - intragenic mapping in, *218,* 218–222, *221*
    - latency in, 212
    - mutations in, 215–218, *216, 217*
    - plaque formation by, 215–218, *216, 217*
    - reproductive cycle of, 209–212, *210–211*
        - lysogenic, 212, *213*
        - lytic, 212, *213,* 215, *216*
    - temperate, 212
    - virulent, 212
- Baculoviruses, *207*
- BAC (bacterial artificial chromosome) vector, 508–509, *509,* 512–513, *513*
- Balanced translocation, 165, *166*
- Balancing selection, *625,* 625–628
- Balbiani, E. G., 172
- Baldness, *98,* 98–99, *99*
- *Bam*RI, *447*
- Barr body, 76, *76*
- Basal transcription apparatus, 310
- Base(s), 232, *232,* 234, 235, *235,* 327–328, *328*
    - RNA editing of, 317–318, *318*
    - wobble, 328, 341, *341*
- Base excision repair, *431,* 431–432, *432*
- Base pairs
    - in DNA, *237,* 237–238
    - mismatch repair of, 433–434
    - in RNA, 240, *240*
- Base sequences
    - in transcription, *300,* 300–301
    - in translation, 327, *328*
- Base substitution, 412, *413*
- Bateson, William, 102, 130, 595–596, 604, 651–652
- Bayesian method, 684
- B DNA, 239, *239*
- Beadle, George, 325–326
- Bees, 72, *73*
- Beijerinck, Martinus, 206
- Benzer, Seymour, 215
- Bertram, Ewart, 76
- Beta (β) sheet, 330–331, *333*
- *bicoid* gene, 125, 591–592, *592, 593*
- Bilateral gynandromorph, 175–176, *176*
- Binary fission, 23–24, *24*
- Binomial expansion equation, 59–60
- Biodegradation, 476
- Bioinformatics, 531, 540–547
    - computer analysis in, 540–541
    - database creation for, 541–542, *542*
    - functional genetic sequence identification for, 542–543, *543*
    - genetic sequence prediction in, *546,* 546–547
    - homologous genes for, 543–545, *544*
- Biolistic gene transfer, 488
- Biological control agents, 476, 488, *489, 490*
- Biological species concept, 676
- Biometrics, 648
- Biotechnology, *444,* 474–477
    - for biological control, 476, 488, *489, 490*
    - for bioremediation, 476–477
    - definition of, 474
    - gene therapy in, 491–495, *492, 493, 494, 495*
    - genetically modified animals in, *477,* 477–483, *478, 479, 480, 481, 482*
    - genetically modified plants in, 488–491, *489, 490, 491*
    - for medicine production, 475, *476,* 482, *482, 483, 575*
    - microorganisms in, 474–477, *475, 476*
    - reproductive cloning in, 483–485, *484, 485*
    - stem cells in, 483–487, *484, 485, 486, 487*
    - transgenic livestock in, 482, *482, 483*
- Biotransformation, 476
- Birds
    - dosage compensation in, 75
    - egg laying by, 624, *624*
    - evolution in, *692,* 692–693
    - feet of, *692,* 692–693
    - sex determination in, 72, *72*
    - sexual dimorphisms in, 99–100, *100*
    - Z chromosome in, 75
- *bithorax* gene, 587–588, *588*
- Bivalent, 30, *31, 32,* 131
- Blackburn, Elizabeth, 286
- BLAST (basic local alignment search tool) program, 544–545, *545*
- Blood group
    - ABO, 96–98, *97*
    - MN, 618–619
- Blotting method
    - Northern, 460, *460*
    - Southern, 458–460, *459*
    - Western, 460–461, *461*
- *BMP4* gene, *692,* 692–693
- Body pattern, 586–587, 589–591, *590, 693,* 693–694
- Bone morphogenetic protein 4, *692,* 692–693
- Botstein, David, 653
- Bottleneck effect, 629, 630
- Boveri, Theodor, 54
- Boycott, Arthur, 123
- Branch migration, 288–290, *289*
- Brassica oleracea, 663, *664*
- BRCA-1 gene, 576–577, *577*
- Breast cancer, 576–577, *577*
- Breeding, selective, *663,* 663–666, *664, 665*
- Brenner, Sydney, 597
- Bridges, Calvin, 81
- Broad-sense heritability, 658
- 5-Bromouracil, 427, *427,* 428, *428*
- Bt corn, 489–490, *491*
- Burke, David, 508
- Bush, Guy, 678
- Butterfly, 7–8, *8*

C

- Cabbage, 663, *664*
- *Caenorhabditis elegans, 13, 518*
    - development of, 597–598, *598*
    - dosage compensation in, 75, *75*
    - mRNA silencing in, *399,* 399–401, *400–401*
- Cain, Arthur, 624
- Calico cat, 76, *76*
- Cancer, 565–577
    - breast, 576–577, *577*
    - cell cycle in, 565–566, *567,* 572–573, *573*
    - chromosome abnormalities in, *568,* 569, *569,* 574, *575*
    - clonal origin of, 565
    - colorectal, 574, *575*
    - DNA repair genes in, 576–577
    - gene amplification in, *568,* 568–569
    - growth of, 565, *565*
    - invasive, 565, *565*
    - loss of heterozygosity in, 576–577, *577*
    - metastatic, 565, *565, 575*
    - missense mutations in, 568, *568*
    - molecular profiling in, 575–576
    - multiple mutations in, 574, *575*
    - oncogenes in, 566, *566, 568,* 568–569
    - *p53* gene in, 571, *572*
    - tumor-suppressor genes in, *272,* 570–572, *576,* 576–577
    - two-hit hypothesis of, 570–571, *571*
    - viruses in, *568,* 569–570, *570*
- *Canis familaris, 518*
- CAP (catabolite activator protein), 358, 359, 365, *365*
- Capping, 316–317, *317*
- *Capsella bursa-pastoris,* 104, *104*
- Capsid, 207, *208*
- Capsomers, 207, *208*
- Carbohydrates, 4
- Carboxyl-terminal end (C-terminus), 329
- Carcinogen, 565
- Carle, Georges, 508
- Caspases, 571
- Castle, William, 651–652
- Cat
    - calico, 76, *76*
    - cloning of, *1,* 3, 484, *485*
    - Manx, 100, *101*
- Catabolic enzyme, 370–371
- Catabolite activator protein (CAP), 358, *359,* 365, *365*
- Catabolite repression, 364–365, *365*
- Catenanes, 280, *280*
- Cavalli-Sforza, Luca, 188, 191
- cDNA, *449,* 449–450, 455–457, *456*
- Cell(s)
    - cancer (*See* Cancer)
    - competent, 199
    - death of, 571, 586, *586*
    - determined, 600
    - development of, 586, *586* (*See also* Developmental genetics)
    - differentiated, 600
    - germ, 10, 21, 34, *35*
    - host, 207, 445
    - nonparental, 131
    - nonrecombinant, 131
    - parental, 131
    - recombinant, 131
    - stem, *485,* 485–487, *486, 487*
    - totipotent, 485, *486,* 602
- Cell adhesion, 587, *587*
- Cell adhesion molecules, 587, *587*
- Cell culture, A–1
- Cell cycle, 24–26, 25, 565–566, *567. See also* Meiosis; Mitosis
    - in cancer, 572–573, *573*
- Cell fate, 595, 600
- Cell-free translation system, *334–335,* 335
- Cell lineage, 597–598, *598*
- Cell plate, 29, *29*
- CentiMorgans (cM), 140
- Central dogma, 299, *300*
- Centrifugation, A-2–A-3
- Centriole, 26, *32*
- Centromere, 24, *26,* 251, *251*
    - cohesion at, 263, *264*
- Centrosome, 26, *32*
- *Cepaea nemoralis,* 624, *625*
- CF gene, 57, 101, 544
- CFTR protein, 57, 544
- Chain termination, 464
- Chalfie, Martin, 4
- Chaperones, 330
- Character, 43, *44*
- Charcot-Marie-Tooth disease, 160
- Chargaff, Erwin, 235
- Chargaff's rule, 235
- Charged tRNA, 340, *340*
- Chase, Martha, 228
- Checkpoint proteins, 572–573, *573*
- Chiasma, 30, *31*
- Chiba, Yukako, 111
- Chicken, foot of, *692,* 692–693
- Chickenpox, 213
- Chimera, 479, *480*
- Chimpanzee, 682–683, *684,* 690, *690*
- ChIP-chip assay, 532–535, *533*
- Chi square test, 60–62, *62,* 135–136
- *Chlamydomonas,* chloroplast inheritance in, 113, *113, 114*
- Chloride transporter, *90*
- Chloroplast, 20
    - extranuclear inheritance in
        - in *Chlamydomonas,* 113, *113*
        - circular chromosomes and, 111
        - maternal inheritance and, 111–112, *112, 113*
        - species variation and, *114,* 114

  - nucleoids of, 111, *111*
  - origin of, 116–117, *117*
- Chorionic villus sampling, *562,* 562–563
- Chou, Peter, 546
- Chromatid(s), 24
  - sister, 24, *26, 27, 32,* 33
- Chromatin, 19
  - closed conformation of, 387
  - interphase, *259,* 259–260
  - nucleosomes of, *254,* 254–257, *255–256, 257,* 386–388, *387*
  - open conformation of, 387
  - remodeling of, 386–388, *387, 390,* 390–391
    - insulator effects on, 394, *395*
- Chromatin immunoprecipitation (ChIP), 532–535, *533*
- Chromatography, A-3–A-4
- Chromocenter, *172,* 173
- Chromosome(s), 4, *5,* 19–23
  - acentric fragment of, 164
  - acrocentric, 156, *157*
  - allodiploid number of, 176–179, *177, 178*
  - alloploid number of, 176, *177*
  - allopolyploid number of, 176, *177*
  - allotetraploid number of, 176–179, *177, 178*
  - aneuploid number of, *168,* 168–171, *169, 170, 171*
  - artificial, 508–509, *509*
  - autopolyploid number of, 176, *177*
  - bacterial (*See* Bacterial chromosomes)
  - balanced translocation of, 165, *166*
  - breakpoint of, 415
  - in cancer, *568,* 569, *569,* 574, *575*
  - circular, 111, *112,* 114–115, *115*
  - cloned gene integration into, *478,* 478–479, *479*
  - colchicine effects on, 179, *179*
  - complete nondisjunction of, 175
  - crossing over of (*See* Crossing over)
  - deficiency of, 158
  - definition of, 19, 24
  - deletion of, 158, *158,* 159, *159*
  - dicentric, 164
  - dicentric bridge of, 164
  - diploid, 10, *11,* 21, *22*
  - duplication of, 158, *158,* 159–161, *160, 161*
  - eukaryotic (*See* Eukaryotic chromosomes)
  - euploid number of, *168,* 168–169
  - G bands of, 156, *157*
  - haplodiploid number of, 171
  - haploid, 10, *11,* 30
  - heterochromatic, 415, *417*
  - homeologous, 177
  - homologous, 10, 21–23, *23*
  - interstitial deletion of, 159, *159*
  - inversion of, 158, *158, 162,* 162–164, *163*
    - gene expression and, 415, *417*
  - karyotype of, 21, *22,* 156, *157*
  - meiotic nondisjunction of, 174–175, *175*
  - metacentric, 156, *157*
  - microscopic examination of, 6, *6,* 155–157, *157*
  - mitotic nondisjunction of, 174, 175–176, *176*
  - numbering of, 156, *157*
  - number of, *168,* 168–179
  - paracentric inversion of, 162, *162, 163,* 164
  - pericentric inversion of, 162, *162, 163,* 164
  - polyploid number of, 179, *179*
  - polytene, 171–173, *172*
  - position effects on, 415, *417*
  - in primates, 690, *690*
  - reciprocal translocation of, *164,* 164–165, *165, 167,* 167–168
  - repetitive sequences of, 159–160, *160*
  - Robertsonian translocation of, 165, *165*
  - segmental duplication of, 161
  - submetacentric, 156, *157*
  - telocentric, 156, *157*
  - terminal deletion of, 159, *159*
  - tetraploid number of, *168,* 169, 171, *172,* 173, *173*
  - tissue number of, 171–173, *172*
  - translocation of, 158, *158, 164,* 164–165, *568,* 569, *569*
    - gene expression and, 415, *417*
  - transpositions in, 517, *519,* 519–525, *520, 521, 522, 523*
  - triploid number of, *168,* 168–169, 173–174, *174*
  - unbalanced translocation of, 165
- Chromosome 1, 690, *690*
- Chromosome 2, 690, *690*
- Chromosome 3, 690, *690*
- Chromosome 4, *2*
- Chromosome 5, 159, *159*
- Chromosome 9, in corn, 136, *137*
- Chromosome 16, 509–510, *510*
- Chromosome looping, 395, *395*
- Chromosome mapping, 500. *See also* Genetic mapping
  - cytogenetic, 500, *500, 501,* 501–503, *502, 510*
  - linkage, 503–506, *505, 506, 510*
  - physical, 500, *500, 507,* 507–512, *510*
  - in situ hybridization for, 501–502, *502*
- Chromosome painting, 502, *502*
- Chromosome territory, 258–259, *259*
- Chromosome theory of inheritance, *53,* 53–56, *54, 55*
- Chromosome walking, 510–511, *511,* 560
- *cis*-acting element, 309, 364
- Cistron, 217
- Clade, 680, *680*
- Cladistics, *681,* 681–682, *683*
- Cladogenesis, 676–678, *677*
  - allopatric, 676–678, *677*
  - parapatric, *677,* 678
  - sympatric, *677,* 678, *679*
- Cladogram, *681,* 681–682, *683*
- *Cla*I, 447
- Cleavage furrow, 29, *29,* 32
- Cloning, 3, *3,* 484, *485. See also* Biotechnology; Gene cloning; Reproductive cloning
- Clownfish, *73,* 73–74
- *CLV3* gene, 603–604, *604*
- Coactivator protein, 383, *383*
- Cockayne syndrome, 433
- Codominance, *89,* 97
- Codon, *300,* 301
  - AUG, 345–346, *346*
  - degeneracy in, 341, *341*
  - nonsense, 327
  - sense, 327
  - start, *300,* 301, 327, 328, *328*
  - stop, *300,* 301, 327, *328*
  - synonymous, 328
  - termination, 327
  - trp, *369,* 369–370, *370*
- Cohesin, 262–263, *264*
- Colchicine, 179, *179*
- Collared flycatcher, 624
- Colony hybridization, *457,* 457–458
- Colorectal cancer, 574, *575*
- Compaction. *See* Eukaryotic chromosomes, compaction of
- Comparative genomics, 517
- Competence factors, 199
- Competence-stimulating peptide, 200
- Competent cell, 199
- Complementarity, 238, *238*
- Complementary DNA (cDNA), *449,* 449–450, 455–457, *456*
- Complementary sequences, 238, *238*
- Complementation, *102,* 103, *103*
- Complementation test, 216–218, *217*
- Complex trait, 646–647, *647. See also* Quantitative genetics
- Composite transposon, 520, *520*
- Computer analysis, in bioinformatics, 540–541
- Concordance, 552
- Condensin, 262–263, *263*
- Conditional lethal allele, 100
- Conditional mutation, 414
- Conglomerate, 630–631
- Conjugation. *See* Bacterial conjugation
- Conjugation bridge, 188, *189*
- Consensus sequence, *303,* 303–304
- Constitutive exon, 397, *397*
- Constitutive gene, 355
- Constitutive heterochromatin, 259
- Contig, *507,* 507–508, *508*
- Copy number variation, *161,* 161–162
- Corepressor molecules, 356, *357*
- Core promoter, 308–309, *309*
- Corn, 136, *137*
  - Bt, 489–490, *491*
  - selective breeding in, 664, *665*
  - synteny groups in, 691, *691*
- Coronaviruses, 214
- Correlation coefficient (*r* value), 650–651, *651*
- Correns, Carl, 94, 111
- Cotransduction, 197–199, *198*
- Cotransformation frequency, 200
- Coumarins, 249–250
- Coupling (*cis* configuration), 140
- Covariance, 650
- cpDNA, 111, *112*
- CpG islands, 391–392, *392*
- Creighton, Harriet, 136
- Cremer, Christoph, 259
- Cremer, Thomas, 259
- Creutzfeldt-Jakob disease, *564*
- Crick, Francis, 235–236, *236*
- Cri-du-chat syndrome, 159, *159*
- Crisanti, Andrea, 4
- Cross-fertilization, 43, *43*
- Crossing over, 30, 51, 287, *288. See also* Recombination, homologous
  - of corn chromosomes, 136, *137*
  - in dihybrid cross, 140–143, *141, 142*
  - in inversion loop, *163,* 163–164
  - linkage and, 131, *131,* 136, *137*
  - in mitosis, 145–147, *146*
  - molecular characteristics of, 287–293, *289, 290, 291, 292*
  - nonallelic homologous recombination and, 159–160, *160*
  - positive interference and, 145
  - in trihybrid cross, 143–144, *144*
  - of X-chromosomes, 131–133, *132, 133, 134*
- Crown gall disease, 476, 488, *489, 490*
- CTC-binding factor, 120, *120*
- Cycle threshold (C_t) method, 455
- Cyclin-dependent protein kinases, 572–573, *573*
- Cyclins, 572–573, *573*
- Cystic fibrosis, 57, *90,* 101, 544, *554*
- Cytogeneticist, 155–156
- Cytogenetic mapping, 500, *500, 501,* 501–503, *502, 510*
- Cytogenetics, 21, *22*
- Cytokinesis, *28,* 29, *29, 32,* 33
- Cytoplasmic inheritance. *See* Extranuclear inheritance
- Cytosine, 5, 232, *232,* 235, *235*
  - methylation of, 239, 391–392, *392*

D

- *Dactylorhiza sambucina,* 90, *90,* 626, *626*
- *Danio rerio, 13*
- Darwin, Charles, 10, 620, 673
- Darwinian fitness, 621
- Database, in bioinformatics, 541–542, *542*
- Davis, Bernard, 188
- DDT (dichlorodiphenyltrichloroethane) resistance, 623, *623*
- Deamination, *421,* 422, *422*
- Degeneracy, 327–328, 340–341, *341*
- Degrees of freedom, 62, *62,* 649
- Delbrück, Max, 419
- Deletion mapping, 220–221, *221*
- *Dendrobates tinctorius,* 8, *9,* 674, *674*
- Deoxyadenosine triphosphate, 232
- Deoxyribonucleic acid. *See* DNA (deoxyribonucleic acid)
- Deoxyribose, 232, *232*
- Depurination, *421,* 421–422, *422*
- De Queiroz, Kevin, 676
- Dermal ridge count, 659–661, *660–661*
- Developmental genetics, 585–605
  - body axes in, 588, *588, 590,* 591–593, *592, 593*

- Developmental genetics, *continued*
  - body pattern generation in, 586–587, 589–591, *590*
  - body segmentation in, 588, *588, 593,* 593–594, *594, 595*
  - cell differentiation in, 588, *588*
  - determination process in, 588, *588*
  - genes in, 587–589, *588*
  - induction in, 587
  - invertebrate, 589–598 (*See also Drosophila melanogaster*)
    - axis of polarity in, *591,* 591–593, *592*
    - in *Caenorhabditis elegans,* 597–598, *598*
    - early embryonic stages of, 589–591, *590*
    - homeotic genes in, 594–597, *596, 597*
    - maternal effect genes in, *591,* 591–593, *592, 593*
    - segmentation genes in, *593,* 593–594, *594, 595*
  - morphogen in, 586–587, *587*
  - overview of, *586,* 586–589, *587, 588, 589*
  - phases of, *588,* 588–589
  - plant, 602–605, *603*
    - homeotic genes in, 604–605, *605*
    - steps in, *603,* 603–604, *604*
  - vertebrate, 599–602
    - homeotic genes in, *599,* 599–600, *600, 601*
    - transcription factor genes in, 600–602, *602*
- Diakinesis, 30, *31*
- Diauxic growth, 364
- Dicentric bridge, 164
- Dicentric chromosome, 164
- Dicer, 402, *402*
- Dideoxy method, 464–465, *465*
- Dideoxyribonucleotides (ddNTPs), 464, *464*
- Differential centrifugation, A-3
- Differentially methylated region (DMR), 395, *395*
- Dihybrid cross, 48–51, *49–50,* 130
  - crossing over in, 140–143, *141, 142*
- Dihybrid testcross, 52
- 5-α-Dihydrotestosterone, 98
- Diplotene, 30, *31*
- Directional selection, 621–623, *622, 623*
- Direct repeats, 519, *520*
- Discontinuous trait, 647
- Discovery science, 15
- Disease. *See also* Cancer; Genetic disease
  - mutations and, 424–425, *425*
  - prion, 563–564, *564*
- Disequilibrium, 619
- Disruptive (diversifying) selection, 624, *624, 625*
- Dizygotic twins, 552
- DNA (deoxyribonucleic acid), 1, *2*
  - attenuation sequence of, 368
  - B, 239, *239*
  - chloroplast (cpDNA), 111, *112*
  - complementary (cDNA), *449,* 449–450, 455–457, *456*
  - copy number variation, *161,* 161–162
  - daughter strands of, *269,* 269–270
  - directionality of, *233,* 234
  - DNase I footprinting of, 462–463, *463*
  - double helix of, 231, *231*
    - alternative types of, *239,* 239–240
    - antiparallel arrangement in, 238
    - ball-and-stick model of, 238, *238*
    - complementary sequences in, 238, *238*
    - discovery of, *234,* 234–236, *235, 236*
    - grooves in, 239
    - space-filling model of, *238,* 239
    - spiral turns of, *238,* 238–239
    - structure of, *237,* 237–240, *238,* 268–269, *269*
  - extrachromosomal, 21
  - extranuclear, 21
  - gel retardation assay of, 462, *462*
  - grooves in, *238*
  - heteroduplex region of, 199, 288
  - information storage by, 5–6
  - insulator, 394–396, *395*
  - length of, 253
  - magnesium hypothesis of, 236
  - methylation of, 391–394, *392, 393, 394*
    - imprinting and, 119–120, *120,* 121
  - mitochondrial (mtDNA), 115, *115*
  - mutation in (*See* Mutation[s])
  - nucleotide structure of, *232,* 232–233, *233*
  - origin of transfer sequence of, 189, *190*
  - phage T2, *229–230,* 229–231
  - promoter of, *300,* 301, *303,* 303–304
  - protein binding to, 239, *462,* 462–463, *463*
  - radial loop domain of, 257–258, *258*
  - recombinant (*See* Recombinant DNA)
  - repair of (*See* DNA repair)
  - replication of, 24, *26*
  - sequence complexity of, 252–253, *253*
  - site-directed mutagenesis of, 466–467, *467*
  - Southern blotting for, 458–460, *459*
  - strand of, 231, *231, 233,* 233–234
  - structure of, 231, *231, 233,* 233–234 (*See also* DNA (deoxyribonucleic acid), double helix of)
  - supercoiling of, *248,* 248–250, *249*
  - template (parental) strands of, *269,* 269–270
  - topoisomers of, 248, *249*
  - T2 phage, *228,* 228–231, *229–230*
  - transferred (T DNA), 488, *489*
  - translesion synthesis of, 437
  - x-ray diffraction of, 234–235, *235*
  - Z, *239,* 239–240
- DnaA box sequences, 274, *274*
- DnaA protein, 274, *274, 276*
- DNA-binding proteins, *462,* 462–463, *463*
- DnaC protein, *276*
- DNA double-strand break, 434–436, *435, 436*
- DNA fingerprinting, 3, 636–637, *637*
- DNA gap repair synthesis, 290–293, *292*
- DNA gyrase, 248–250, *250, 275,* 276, *276*
- DNA helicase, *274,* 274–275, *276*
- DNA library, 455–461
  - cDNA, *456,* 457
  - construction of, 455–457, *456*
  - genomic, 55, *456*
  - in physical mapping, 507
  - screening of, *457,* 457–458
- DNA ligase, *276,* 278, *446,* 447
- DNA methylation, 239, 391–394, *392, 393, 394*
  - de novo, 392–393, *394*
  - imprinting and, 119–120, *120, 121*
  - maintenance, 392–393, *394*
- DNA methyltransferase, 391–392, *392*
- DNA microarray, *532,* 532–535, *534*
- DNA mismatch repair, *431,* 433–434, *435*
  - gene conversion by, 291–293, *292*
- DNA *N*-glycosylases, 431, *432*
- DNA polymerase(s)
  - bacterial, *276, 279, 281,* 281–282
  - in replication, *275,* 276–277, *277*
  - eukaryotic, *284,* 284–285
  - lesion-replicating, 284, *284,* 437
  - proofreading function of, *282,* 282–283
- DNA polymerase I, *276,* 276–277
- DNA polymerase III, *275, 276,* 276–277, *277, 279, 281,* 281–282
- DNA profiling, 636–637, *637*
- DNA repair, 430–437
  - alkyltransferase in, 430–431, *431*
  - base excision in, *431,* 431–432, *432*
  - direct, 430–431, *431*
  - homologous recombination in, *431,* 434–436, *435*
  - mismatch repair in, *431,* 433–434, *435*
  - nonhomologous end joining in, *431,* 434–436, *436*
  - nucleotide excision in, *431,* 432–433, *433*
  - photolyase in, 430, *431*
  - transcription-coupled, 436–437, *437*
  - vs. translesion synthesis, 437
- DNA replication, 268–293
  - bacterial, 272–283
    - accuracy of, *282,* 282–283
    - bidirectional, *274,* 275
    - catenanes in, 280, *280*
    - chemistry of, *281,* 281–283
    - dimeric DNA polymerase in, 278–279
    - DnaA box sequences in, 274, *274*
    - DNA helicase in, *274,* 274–275
    - DNA ligase in, *276,* 278
    - DNA polymerase in, *275,* 276–277, *277, 281,* 281–282
    - double helix unwinding in, 276
    - lagging strand in, *275,* 276, 277–278, *278, 279*
    - leading strand in, *275,* 276, 277–278, *278, 279*
    - origin of replication in, 272–275, *273, 274*
    - primase in, 276, *276,* 278–279, *279*
    - primosome in, 278–279, *279*
    - proofreading in, *282,* 282–283
    - proteins in, *275,* 275–277, *276, 277*
    - replication forks in, 272–277, *273, 275*
    - replisome in, 278–279, *279*
    - RNA primer synthesis in, 276
    - single-strand binding proteins in, 276
    - termination of, 279–281, *280*
    - ter sequences in, 279–280, *280*
    - topoisomerase II in, *275,* 276, *276*
  - conservative model of, 270, *270*
  - daughter strands in, *269,* 269–270
  - dispersive model of, 270, *270*
  - eukaryotic, 283–287
    - ARS elements in, 283
    - DNA polymerases in, *284,* 284–285
    - flap endonuclease in, 285, *285*
    - MCM helicase in, 284
    - origin recognition complex in, 284
    - origins of replication in, *283,* 283–284, *284*
    - prereplication complex in, 284
    - RNA primer removal in, 285, *285*
    - telomerase in, 285–286, *286, 287*
  - models of, *270,* 270–272
  - semiconservative model of, *270,* 270–272
  - structural basis of, 268–272, *269, 270*
  - template (parental) strands in, *269,* 269–270
- DNA replication licensing, 284
- DNase, 228
- DNase I, chromatin cutting with, 255–256, *255–256*
- DNase I footprinting, 462–463, *463*
- DNA sequencing, 464–467
  - automated, 464, *465*
  - dideoxy method of, *464,* 464–466, *465*
  - high-throughput, 514–515
  - next-generation technologies for, 514–517, *515*
  - pyrosequencing for, 515, *516*
  - shotgun, 512–514, *513*
- DNA translocases, 387
- DNA uptake signal sequences, 200
- Dobzhansky, Theodosius, 675
- Dogs
  - muscular dystrophy in, *82,* 82–83
  - selective breeding of, 663, *663*
- Dolly (cloned sheep), 3, *3,* 483–485, *484*
- Dominance, *91,* 91
  - autosomal, 554–556, *555*
  - codominance, 97
  - incomplete, *89, 94,* 94–95
  - overdominance, *95,* 95–96, *96*
  - X-linked, 556, *557*
- Dominant allele, 91
- Dominant inheritance, *90,* 90–92, *91*
  - autosomal, 554–556, *555*
  - X-linked, 556, *557*
- Dominant-negative mutation, 91, 555–556
- Dorsoventral axis, *590,* 591
- Dosage compensation, *75,* 75–78
- Double helix, 231, *231*
  - alternative types of, *239,* 239–240
  - discovery of, *234,* 234–236, *235, 236*
  - structure of, *237,* 237–240, *238,* 268–269, *269*
- Double-strand break model, *290,* 290–291
- Down regulation, 383
- Down syndrome, 9, *9, 170*
  - familial, 165, *166*
  - maternal age and, 170–171, *171*
- *Drosophila melanogaster, 13, 518*
  - aneuploid, *168,* 169
  - *Antennapedia* mutation in, 596, *596*
  - *bicoid* gene in, 125, 591–592, *592, 593*
  - bilateral gynandromorph of, 175–176, *176*
  - *bithorax* gene in, 587–588, *588*

  - body color in, 61–62, 131–136, *132, 133, 134, 139,* 139–144, *141, 144, 146,* 146–147
  - bristle morphology in, *139,* 139–140, *146,* 146–147
  - chromosome number in, *168,* 168–169
  - development of, 589–598
    - axis of polarity in, *591,* 591–593, *592*
    - early embryonic stages of, 589–591, *590*
    - homeotic genes in, 594–597, *596, 597*
    - maternal effect genes in, *591,* 591–593, *592, 593*
    - segmentation genes in, *593,* 593–594, *594, 595*
  - dihybrid cross in, 140–143, *141*
  - diploid, 168, *168*
  - dosage compensation in, 75, *75*
  - evolution of, *689,* 689–690
  - eye color in, 75, *75,* 79–80, *80–81,* 100, 131–136, *132, 133, 134,* 140–144, *141, 144,* 415, *417*
  - eye of, 93, *93,* 694–695, *695*
  - genetic linkage map of, 137–138, *138*
  - Hedgehog in, 587
  - longevity of, 657–658, *658*
  - maternal effect genes in, 125
  - mitotic recombination in, 145–147, *146*
  - monosomic, *168,* 169
  - P element in, 522, *524,* 524–525
  - polyploid, *168,* 169
  - polytene chromosomes in, 171–173, *172*
  - salivary gland cell chromosomes of, 171–173, *172*
  - selective breeding in, 664, *665*
  - sex chromosomes in, 175–176, *176*
  - testcross in, *139,* 139–140
  - tetraploid, *168,* 169
  - transposable elements in, 522, *524,* 524–525
  - trihybrid cross in, 131–132, *132, 133,* 143–144, *144*
  - triploid, *168,* 168–169
  - trisomic, *168,* 169
  - wing shape in, 61–62, 131–136, *132, 133, 134,* 143–144, *144*
  - X-linked traits in, 131–134, *132, 133*
- Duchenne muscular dystrophy, 82, *82, 556*
- Duck, foot of, *692,* 692–693
- Duplication, 158, *158,* 159–161, *160, 161*
- Dwarfism, 630

E

- E6-AP, 121
- Ecological species concept, 676
- *Eco*RI, *446,* 446–447, *447,* 504, *504*
- Ecotype, 674
- Edward syndrome, *170*
- Egg cell, 10, 21, 34, *35,* 485, *486*
- Ehler-Danlos syndrome, 581
- Elderflower orchid, 90, *90,* 626, *626*
- Electron microscope, *208,* 207
- Electrophoresis, A-3–A-4
- Electroporation, 488
- Ellis-van Creveld syndrome, 630
- Elongation stage
  - in transcription, *301,* 301–302
  - in translation, 344, *344, 346,* 346–347
- Embryogenesis
  - *Drosophila,* 589–598
    - axis of polarity in, *591,* 591–593, *592*
    - early stages of, 589–591, *590*
    - homeotic genes in, 594–597, *596, 597*
    - maternal effect genes in, *591,* 591–593, *592, 593*
    - segmentation genes in, *593,* 593–594, *594, 595*
  - maternal effect genes in, 123–125, *124*
  - morphogenic gradients in, 587, *587*
  - plant, *603,* 603–604
  - vertebrate, *599,* 599–602, *600, 601*
- Embryonic germ cells, 485
- Embryonic stem cells, 485, *486*
- Embryo sac, 36, *36*
- Empirical law, 44
- Endonuclease, 312, *314*
- Endoplasmic reticulum, 20, *20*
- Endopolyploidy, 171–173, *172*
- Endosperm, 36
- Endosymbiosis, *116,* 116–117
- Endosymbiosis theory, 116–117, *117*
- Enhancer, 308–309, 380, *381*
  - bidirectional, 383
  - insulator effects on, 394, 395, *395*
  - splicing, 398, *398*
  - transcription factor recognition of, 382–383, *383*
- *Entamoeba histolytica, 518*
- Environment
  - gene expression and, 92–93, *93*
  - genotype interactions with, 657–658, *658*
  - polygenic inheritance and, 652–653, *653, 654*
  - sex determination and, *73,* 73–74
  - traits and, 9–10, *10*
- Enzyme(s), 5. *See also specific enzymes*
  - allosteric, 372, *372*
  - catabolic, 370–371
  - feedback inhibition of, *372,* 372–373
  - *Neurospora,* 325–327, *326*
  - processive, 282
  - regulation of (*See* Gene regulation)
  - repressible, 370–371
  - synthesis of, *325,* 325–327, *326* (*See also* Translation)
- Enzyme adaptation, 358
- Epidermal growth factor, 566, *567*
- Epigenetic inheritance, 118–122
  - DNA methylation in, 119–120, *120,* 121
  - genomic imprinting in, *118,* 118–119, *119,* 121–122, *122*
  - human disease and, 120–121, *122*
- Episomes, 189, 212–213
- Epistasis, *102,* 102–103, *103*
- Epitope, 460
- Epstein-Barr virus, *207*
- Equilibrium, Hardy-Weinberg, 617–619, *618*
- Equilibrium density centrifugation, A-3
- *Equus,* 11, *12*
- Error-prone replication, 437
- *Escherichia coli,* 12, *13*
  - base mismatch repair in, 433–434, *435*
  - DNA supercoiling in, 248
  - genetic mapping in, 193–195, *194, 195*
  - genetic transfer in, *187,* 187–189, *188, 189, 191,* 191–192, *192*
  - genome of, *518*
  - *Hfr* strains of, *191,* 191–196, *192*
  - homologous recombination in, 291, *291*
  - insulin production in, 1, 475, *476*
  - *lac* operon of, 358–367, *360, 361, 362–363, 364*
  - *lacY* gene in, 543, *544*
  - *micF* gene in, 372, *372*
  - nucleotide excision repair in, 433, *433*
  - ompF protein in, 372
  - phage T2 infection of, *228,* 228–231, *229–230*
  - phage T4 infection of, 215–218, *216, 217*
  - promoter in, *303,* 303–304
  - riboswitch in, 373–374, *374*
  - SOS response in, 437
  - 16S rRNA from, 543, *546*
  - thiamin pyrophosphate riboswitch in, 373–374, *374*
  - transcription-coupled DNA repair in, 436, *437*
  - transcription in, 302–303
  - *trp* operon of, 367–371, *368, 369*
- *Essay on the Principle of Population, An* (Malthus), 673
- Essential gene, 100
- Estradiol, in clownfish, 74
- Ethyl methanesulfonate, 427, *427*
- Euchromatin, 259, *259*
- Eukaryote(s), 20, *20*
  - cell cycle of, 24–26, *25,* 565–566, *567*
  - chromosomes of (*See* Eukaryotic chromosomes)
  - DNA replication in (*See* DNA replication, eukaryotic)
  - gene regulation in (*See* Gene regulation, eukaryotic)
  - supergroups of, 684, *685*
  - transcription in (*See* Transcription, eukaryotic)
  - translation in (*See* Translation, eukaryotic)
- Eukaryotic chromosomes, 21–23, *22, 251,* 251–252
  - *Aal* sequence of, 253
  - during cell division, 260–264, *261, 262, 263, 264*
  - centromere of, 251, *251*
  - chromatin of, *254,* 254–255
  - chromosome territory of, 258, *259*
  - compaction of, 253–263
    - cohesin in, 262–263, *264*
    - condensin in, 262–263, *263*
    - interphase, *259,* 259–260
    - metaphase, 260–261, *261, 262*
    - nucleosome in, *254,* 254–257, *255–256, 257, 261*
    - protein scaffold in, 260, *261*
    - radial loop domains in, 257–258, *258, 261*
    - SMC proteins in, 262, *262*
    - 30-nm fiber in, 256–257, *257, 261*
  - highly repetitive sequences of, 253
  - introns of, 252
  - kinetochore of, 251, *251*
  - matrix-attachment regions of, 258, *258*
  - metaphase, 260–263, *261, 262*
  - moderately repetitive sequences of, 252–253, *253*
  - in nondividing cells, 253–260
  - nucleosomes of, *254,* 254–257, *255–256, 257*
  - origin of replication of, 251, *251*
  - radial loop domains of, 257–260, *258, 259*
  - retroelement of, 253
  - sequence complexity of, 252, *253*
  - tandem array of, 253
  - telomeres of, 251, *251*
  - transposable elements of, 253
- Euploidy, *168,* 168–169
- variations in, 171–174, *172, 173*
- E-value, 545
- *Even-skipped* gene, 591, *592*
- Evolution, 10–11, *12,* 672–696. *See also* Microevolution
  - chromosome changes in, 690, *690*
  - Darwin on, 673
  - definition of, 672
  - endosymbiosis theory in, 116–117, *117*
  - gene duplications and, *160,* 160–161
  - genetic variation and, 14
  - *gremlin* gene in, 692–693, *693*
  - Hardy-Weinberg equation for, 618–619
  - horizontal gene transfer in, 684–685, *686*
  - *Hox* genes in, *693,* 693–694
  - molecular (*See* Molecular evolution)
  - natural selection and (*See* Natural selection)
  - neutral theory of, 688
  - origin of species and, 673–679, *674, 675, 677*
  - *Pax6* gene in, 694–695, *695*
  - phylogenetic trees in, 679–685, *680, 681, 683, 684, 685, 686*
  - transposable elements and, 522, 524–525, *525*
- Evolutionary developmental biology, *692,* 692–695, *693, 695*
- Evolutionary lineage concept, 676
- Exit (E) site, 343, *343*
- Exon, 312
  - alternative, 397, *397*
  - constitutive, 397, *397*
- Exon shuffling, 522, 635, *635*
- Exon skipping, 397–398, *398*
- Exonuclease, 312, *314*
- Expressivity, 91–92
- Extranuclear inheritance
  - chloroplast, 110–114
    - in *Chlamydomonas,* 113, *113, 114*
    - circular chromosomes in, 111, *112*
    - maternal inheritance and, 111–112, *112, 113*
    - species variation in, *114,* 114
  - mitochondrial, 114–116
    - maternal inheritance and, *115,* 115–116
    - mutation and, 116, *116*
    - species variation in, *115,* 115–116
- Eye
  - of *Drosophila melanogaster,* 93, *93,* 694–695, *695*
  - evolution of, 694–695, *695*
- Eye color, in *Drosophila melanogaster,* 75, *75,* 79–80, *80–81,* 100, 131–136, *132, 133, 134,* 140–144, *141, 144,* 415, *417*

F

Fabry disease, 581
Factor IX, 482, *482*
Faculative heterochromatin, 259–260
Familial Down syndrome, 165, *166*
Familial fatal insomnia, *564*, 582
Familial hypercholesterolemia, *555*, 579
Fasman, Gerald, 546
Fearon, Eric, 574
Feder, Jeffrey, 678
Feedback inhibition, *372*, 372–373
Felis concolor coryi, 633
Fertility
    allodiploid, 176–179, *178*
    allotetraploid, 177–178, *178*
Fertilization
    cross, 43, *43*
    plant, 36, *42*, 42–43
Fetus, genetic testing of, *562*, 562–563
F factor, 188–189, *189, 190, 191*, 191–192, *192*
F_1 generation, 44, *45–46*
F_2 generation, 44, *45–46*
Fibrillin-1 gene, 556
Ficedula albicollis, 624
Finches, 615, *615*, 626–628, *627*
Fine-structure mapping, *218*, 218–222, *221*
Fingerprint, 659–661, *660–661*
Fire, Andrew, 400
Fitness, Darwinian, 621
Flemming, Walther, 54
Florida panther, 633
Flowering plants. *See also* Plant[s]
    development of, 602–605, *603, 604, 605*
    reproduction in, 35–36, *36, 42*, 42–43
    sex determination in, 74, *74*
Fluctuation test, *419*, 419–420
Fluorescence in situ hybridization, 501–502, *502*
Folding
    polypeptide, 330–331, *333*
    RNA, 240–241, *241*, 339
Forensics, 637
Forked-line method, 51
Founder, 559
Founder effect, 630, *630*, 677–678
Four-o'clock plant, 94, *94*, 111, *112, 113*
Foxtail millet, 691, *691*
Fraction, liquid, A-3
Fragile X syndromes, 424, *425*
Frameshift mutation, 333–334, *334*, 412, *413*
Franklin, Rosalind, 234–235, *235*, 236
Frequency distribution, 647, *648*, 648–649
Frogs, 8, *9*, 171, *172*, 674, *674*
FtsZ protein, 24, *24*
Functional genomics, 531, 532–535, *533, 534*
Functional protein microarray, 539

G

Gain-of-function mutation, 91, 555, 566, *566*
Galápagos finches, 626–628, *627*
Galas, David, 462
galE gene, 196, *196*
Galeopsis tetrahit, 678, *679*
Gallo, Robert, 566
Galton, Francis, 648, 651
Gametes, 10, 21
Gametogenesis, 34
Gametophyte, 35–36, *36*
Gap genes, 593–594, *595*
Garrod, Archibald, 325
Gaucher disease, 581
G bands, 156, *157*
Gearhart, John, 486
Gehring, Walter, 694–695
Gel electrophoresis, *537*, 537–538, A-3–A-4
Gellert, Martin, 248
Gel retardation assay, 462, *462*
Gene(s), 2
    chloroplast, 111, *112*
    constitutive, 355
    copy number variation in, *161*, 161–162
    definition of, 4, 46, 299
    developmental, *692*, 692–695, *693, 695*
    disease-causing, 558–563, *559, 560*
    distance between, 132–133, *134, 139*, 139–140, 195, *196*
    duplication of, 159–161, *160, 161*
    essential, 100
    hemizygous, 79
    heterozygous, 23, *23*, 46–47
    homeotic, 587, 594–597, *596*
    homologous, 160, *160*, 543–545, *544*, 599
    homozygous, 22–23, *23*, 46
    housekeeping, 392
    inducible, 356, *357*
    interactions of, *102*, 102–105, *103, 104*
    linkage of (*See* Linkage)
    maternal effect, *591*, 591–593, *592, 595*
    mitochondrial, 114–115, *115*
    molecular expression of, *3, 4*, 4–7, 52
    monomorphic, 615
    mutant, 8–9, 14 (*See also* Mutation[s])
    nonessential, 100
    nuclear, 111
    nucleosome-free region of, *389*, 389–390
    pleiotrophic, 101–102
    polymorphic, 615–616, *616*
    position effect on, 162
    pseudoautosomal, 78–79, *79*
    repressible, 356, *357*
    selectable marker, 445
    sex-linked, 78–79, *79*
        transmission of, 79–83, *80–81, 82*
    structural, 299, 324
    synteny of, 130
    tissue-specific, 392
    tumor-suppressor, *272*, 570–572, *576*, 576–577
    X-linked, 81
    zygotic, 594
Gene addition, *478*, 478–479, *479*
Gene amplification, *568*, 568–569
Gene chip, 532, *532*
Gene cloning, 444–450. *See also* Biotechnology
    applications of, 444, *444*
    chromosomal DNA in, 444–445
    chromosome integration with, *478*, 478–479, *479*
    definition of, 444
    in DNA library construction, 455–461, *456*
    DNA ligase in, *446*, 447
    mRNA for, *449*, 449–450
    procedure for, 447–449, *448*
    restriction enzymes in, 445–447, *446, 447*
    reverse transcriptase in, *449*, 449–450
    vector DNA in, 444–445, *445*, 447–449, *448*
Gene conversion, 288, 291–293
    by DNA gap repair, 291–293, *292*
    by DNA mismatch repair, 291–293, *292*
Gene duplication, 159–161, *160, 161*
Gene expression, 6, *6*, 299. *See also* RNA modification; Transcription; Translation
    base sequence functions in, *300*, 300–301
    chromosome rearrangements and, 415, *417*
    colinearity of, 312
    in development, 593–597, *595, 596*
    environmental effects on, 92–93, *93*
    monoallelic, 118
    mutation effects on (*See* Mutation[s])
    position effect in, 415, *417*
    regulation of (*See* Gene regulation)
Gene family, 160–161, *161*, 458, 544
Gene flow, 631
Gene knockin, 481–482, *482*
Gene knockout, 103–105, 478, *481*, 481–482, 535, 599–600
Gene pool, 614–615
General lineage concept, 676
General transcription factors, *309*, 309–310, *310*, 380
Gene redundancy, *102*, 103–104, *104*, 481
Gene regulation
    bacterial, 355–374, *356, 357*
        activator protein in, 364–365, *365*
        attenuation in, *368*, 368–370, *369*
        posttranslational, 356, *356, 372*, 372–373
        repressor protein in, 360–362, *361, 362–363, 364*, 368, *368*
        riboswitches in, *373*, 373–374, *374*
        translational, 371–374, *372, 374*
    cis-effect, 364
    eukaryotic, 379–404, *380*
        alternative splicing in, *396*, 396–398, *397, 398*
        chromatin remodeling in, 386–388, *387, 390*, 390–391
        combinatorial control in, 380–381
        DNA methylation in, 391–394, *392, 393, 394*
        histone variation in, *388*, 388–389, *389*
        insulators in, 394–396, *395*
        model of, *390*, 390–391
        mRNA stability in, *396*, 398–399, *399*
        RNA interference in, *396*, 401–402, *402*
        transcription factors in, *380*, 380–386, *381* (*See also* Transcription factor(s), regulatory)
        translational, *396*, 403, *403, 404*
    trans-effect, 364
Gene replacement, *478*, 478–479
    in mouse, 479–481, *480*
Gene therapy, 491–495, *492*
    definition of, 491
    gene transfer in, 492, *493–494*
Genetically modified animals, *477*, 477–483, *478, 479, 480, 481, 482*
Genetically modified organisms, 474
    biological control by, *475*, 476
    bioremediation by, *475*, 476–477
    mammalian, *477*, 477–483
    medicine production by, 475, *475, 476*
Genetically modified plants, 488–491, *489, 490, 491*
Genetic code, 5, 327–328, *328*
    degeneracy of, 327–328, 340–341, *341*
    exceptions to, 329, *329*
    experimental determination of, 333–338, *334, 335–336, 337*
    triplets of, 333–334, *334*
Genetic cross, 13, 41–43
    dihybrid, 48–51, *49–50*, 130, 140–143, *141, 142*
    monohybrid, 44–46, *45*
    reciprocal, *82*, 82–83, 123
    trihybrid, 130, 131–132, *132, 133*, 143–144, *144*
Genetic disease, 9, *9*, 551–577. *See also* Cancer
    age of onset of, 552
    allele identification in, 558–563, *559, 560*
    aneuploidy and, *170*, 170–171
    autosomal dominant inheritance of, 554–556, *555*
    autosomal recessive inheritance of, 553–554, *554*
    chromosome deletion and, 159, *159*
    chromosome duplication and, 160
    chromosome inversion and, 162
    founder effect and, 630
    gene therapy in, 491–495, *492, 493–494*
    haplotype association studies in, *559*, 559–561, *560*
    imprinting in, 120–121, *122*
    inbreeding and, 633
    inheritance of, 552–558, *553*
    locus heterogeneity in, 556, 558
    mitochondrial, 116, *116*
    nucleotide excision repair defects and, 433
    pedigree analysis in, 57, 553–556, *554, 555, 557*, 579–582
    prenatal testing for, *562*, 562–563
    Robertsonian translocation and, 165
    testing for, *561*, 561–563, *562*
    in twins, 552
    X-linked dominant inheritance of, 556, *557*
    X-linked recessive inheritance of, 556, *556*
Genetic drift, *620*, 628–630, *629, 630*
Genetic mapping, 500, *500*
    intergenic, 137–145, 218
        cotransduction for, 197–199, *198*

        - cotransformation frequency for, 200
        - dihybrid cross in, 140–143, *141*
        - in *Drosophila melanogaster,* 137–138, *138*
        - in *Escherichia coli,* 193–195, *194, 195*
        - interrupted mating technique in, 193–195, *194, 195*
        - purpose of, 137
        - testcross in, *139,* 139–140, 142, *142*
        - trihybrid cross in, 143–144, *144*
    - intragenic, *218,* 218–222, *221*
        - coinfection experiment in, 218, *218*
        - deletion mapping in, 220–221, *221*
        - hot spots in, 221, *221*
        - map distance in, 220
        - molecular implications of, 222
        - strategy of, 218–220, *219*
- Genetic mosaic, 175–176, *176,* 417
- Genetic polymorphism, 90, *90,* 505–506, *506,* 615–616, *616*
    - balancing selection and, *625,* 625–628
    - disruptive selection and, 624, *624, 625*
- Genetic recombination, 51, 131, 199, 289. *See also* Crossing over; Independent assortment; Recombination, homologous
- Genetics, 1–4
    - central dogma of, 299, *300*
    - definition of, 4
    - developmental (*See* Developmental genetics)
    - evolutionary (*See* Evolution)
    - fields of, 12–15
    - four levels of, 7, *8*
    - immune (*See* Immunogenetics)
    - molecular, 14 (*See also* Molecular evolution)
    - population (*See* Population genetics)
    - quantitative (*See* Quantitative genetics)
    - transmission, 13–14
- Genetic screening, 561, 562
- Genetic testing, *561,* 561–563, *562*
- Genetic transfer, bacterial, *186,* 186–187
    - conjugation, 186, *186, 187,* 187–190, *188, 189, 190*
    - transduction, 186, *186,* 196–199, *197, 198*
    - transformation, 186, *186,* 199–200, *200*
- Genetic variation, 8–9, *9,* 14, 155, 673
    - haplotype, 558–559, *559*
    - population (*See* Population genetics)
- Genome, 1, *2,* 246, 499
    - bacterial, 246, 517, *518,* 542
    - eukaryotic, *252,* 252–253, *253*
    - mitochondrial, 246
    - nuclear, 246
    - size of, 252, *252*
    - viral, 209, 212–213
- Genome maintenance, 572–574, *573*
- Genome-sequencing projects, 512–517, *513, 518*
- Genomic imprinting. *See* Imprinting
- Genomics, 499
    - bioinformatics in, 540–547, *543, 544, 546*
    - chromatin immunoprecipitation in, 532–535, *534*
    - comparative, 517
    - cytogenetic mapping in, *501,* 501–503, *502*
    - DNA microarray in, 532–535, *533, 534*
    - functional, 500, 531, 532–535, *533, 534*
    - gene knockout collection in, 535
    - genome-sequencing projects in, 512–517, *513, 518*
    - linkage mapping in, *503,* 503–506, *504, 505,* 653, *655,* 655–656
    - physical mapping in, *507,* 507–512, *508, 509, 510, 511*
    - proteomics and, 535–540, *537, 538, 539*
- Genotype, 46. *See also* Phenotype
    - Darwinian fitness of, 621
    - environment interaction with, 657–658, *658*
    - halpoid, 558–559, *559*
- Genotype-environment association, 658
- Genotype-environment interaction, 657–658, *658*
- Genotype frequency, 616–617
- *Geospiza magnirostris,* 615, *615*
- Germ cells, 10, 21, 34, *35*
- Germ-line mutation, 417–418, *418*
- Gerstmann-Straüssler-Scheinker disease, *564*
- *GFP* gene, *3,* 3–4
- Gierer, Alfred, 231
- Gilbert, Walter, 464
- Globin genes, 160–161, *161,* 616, *616,* 625, 687, *687, 688*
- GloFish, *478,* 478–479
- Glucocorticoid receptors, 385–386, *386*
- Glucose, in *lac* operon regulation, 364–365, *365*
- Golgi body, 20, *20*
- *Gooseberry* gene, 592, *592*
- Gorilla, 682–683, *684,* 690, *690*
- G phase, of cell cycle, 24, 25, *25*
- Grant, Peter, 626–627
- Grant, Rosemary, 626–627
- Green fluorescent protein, *3,* 3–4
- Greider, Carol, 286
- *Gremlin* gene, *692,* 692–693
- Griffith, Frederick, 226–227
- Griffith, John Stanley, 563
- Groudine, Mark, 566
- Ground finch, 615, *615,* 626–628, *627*
- Growth factors, 566, *567*
- Guanine, 5, 232, *232,* 235, *235*
    - oxidation of, 424, *424*
- Guanosine triphosphate (GTP), 232
- Gustafsson, Lars, 624
- Gynandromorph, bilateral, 175–176, *176*

H

- *Haemophilus influenzae,* 512
- Haig, David, 121–122
- Hairpin, with trinucleotide repeat sequence, 425, *426*
- Haplodiploid system, 72, *73*
- Haploid chromosomes, 10, *11,* 30
- Haploinsufficiency, 91, 555
- Haplotype, 558–560, *559*
- Haplotype association studies, *559,* 559–561, *560*
- Hardy, Godfrey Harold, 617
- Hardy-Weinberg equation, 617–619, *618*
- Hawaiian happy-face spider, 615, *616*
- Hayes, William, 188, 192
- *Hb* genes, 160–161, *161,* 616, *616,* 625, 687, *687, 688*
- Hedgehog, 587
- Height, frequency distribution of, 647, *648*
- Heitz, Emil, 259
- *Helicobacter pylori, 518*
- Helix-loop-helix motif, 381, *382,* 601
- Helix-turn-helix motif, 304, *304,* 381, *382*
- Hematopoietic stem cells, 485–487, *487*
- Hemizygosity, 79
- Hemoglobin, *95,* 95–96, 160–161, *161,* 387, 616, *616,* 687, *687, 688*
- Hemophilia, 162, *556,* 556–558, *557*
- Hemp nettle, 678, *679*
- Heritability, 656–663, *660–661, 662*
    - broad-sense, 658
    - definition of, 656–657, 662, *662*
    - of dermal ridge count, 659–661, *660–661*
    - environment and, 662
    - narrow-sense, 659
    - realized, 665–666
    - selective breeding estimate of, 665–666
- Hermaphrodite, protandrous, *73,* 73–74
- Hernig, Willi, 681
- Herpes simplex virus type I, 213
- Herpes simplex virus type II, *207,* 213
- Hershey, Alfred, 228
- Hertwig, Oscar, 54
- Heterochromatin, 259, *259*
    - constitutive, 259
    - faculative, 259–260
    - gene expression and, 415, *417*
- Heterodimer, 382
- Heteroduplex, 288
- Heterogametic sex, 72
- Heterogamous species, 34
- Heteroplasmy, 111–112
- Heterozygosity, loss of, 576–577, *577*
- Heterozygote
    - Hardy-Weinberg estimation of, 619
    - inversion, *163,* 163–164
    - protein production in, 91, *91,* 96, *96*
- Heterozygote advantage, *95,* 95–96, 625–628
- Hexosaminidase, *90,* 553–554, 561
- *Hfr* strains, *191,* 191–196, *192*
- *H19* gene, 395, *395*
- Hierarchical shotgun sequencing, 512–513, *513*
- Highly repetitive sequence, 253
- High stringency condition, 460
- High-throughput sequencing, 514–515
- Hippocrates, 40
- Histone acetyltransferase, 388–389, *389, 390*
- Histone cenH3, 388
- Histone code hypothesis, 389
- Histone deacetylase, 394, *395*
- Histone H2A.X, 388
- Histone macroH2A, 388
- Histone proteins, *254,* 254–255
    - modification of, 388–389, *389, 390,* 390–391
    - structure of, *254,* 254–255, 388
    - variants of, 388, *388,* 391
- Histone spH2B, 388
- HIV. *See* Human immunodeficiency virus (HIV)
- *HMS Beagle,* 673
- H1N1 virus, 213–214, *214*
- Hoagland, Mahlon, 338
- Holliday, Robin, 288, 392
- Holliday junction, 288
- Holliday model, 288–290, *289*
- Holt, Sarah, 659–660
- Homeobox, 597, *597*
- Homeodomain, 597
- Homeotic genes
    - invertebrate, 594–597, *596, 599*
    - plant, 604–605, *605*
    - vertebrate, *599,* 599–600, *600*
- Homodimer, 96, *96,* 382, 384, *385*
- Homogametic sex, 72
- Homolog, 10, 21
- Homologous attribute, 680
- Homologous recombination. *See* Recombination, homologous
- Homologous recombination repair, *431,* 434–436, *435*
- Homology, 544, 679–680
- Homology-based modeling, 546–547
- *Homo sapiens, 518,* 682–683, *684,* 690, *690*
- Homozygote, protein production in, 91, *91*
- Honey bees, 72, *73*
- Horizontal gene transfer, 636, *636,* 684–685, *686*
- Hormones, in clownfish, 73–74
- Horowitz, Rachel, 257
- Horse, 11, *12*
- Hot spots, 420, 424–425, *425, 426*
- Housekeeping genes, 392
- *HoxC-6* gene, 600, *601*
- *Hox* genes, *693,* 693–694
    - invertebrate, 596, *596, 693*
    - vertebrate, 599–600, *600, 601, 693*
- Human Genome Project, 1, *2,* 514
- Human immunodeficiency virus (HIV), *207,* 214, *214*
    - latency in, 212
    - reproductive cycle of, 209–212, *210–211*
- Hungerford, David, 569
- Huntington disease, 424, *425,* 554–556, *555*
    - haplotype association study in, *560,* 560–561
- Hurler disease, 581
- Hybrid, 41, 678
- Hybrid dysgenesis, 525
- Hybridization, 41–44, *42–44. See also* Genetic cross
- Hybrid zone, 678
- Hydrogen bone, *237,* 237–238, *238*
- Hypercholesterolemia, familial, *555,* 579
- Hypermutation, somatic, 607
- Hypothesis testing, 15, 60–62, *62*
- Hypoxanthine-guanine phosphoribosyl transferase, *90*

I

- *Igf1* gene, 663
- *Igf2* gene, 395, *395*

Ilex opaca, 74, *74*
Illegitimate genetic recombination, 199
Immunogenetics, 605–607
  imprecise end joining in, 607
  site-specific recombination in, *606,* 606–607, *607*
  somatic hypermutation in, 607
Immunoglobulins, 605–607, *606, 607*
Imprinting, 118–122, *120, 121,* 393, *394*
  biological significance of, 121–122
  cellular stages of, 119, *119*
  definition of, 118
  human disease and, 120–121, *122*
  monoallelic expression and, *118,* 118–119, *119*
  in *Sciara coprophila,* 119
Imprinting control region (ICR), 120, *120, 121,* 395, *395*
Inbreeding, 631–633, *632,* 657
Inbreeding coefficient, 632
Inbreeding depression, 633
Incomplete dominance, *94,* 94–95
Incomplete penetrance, 91–92, *92*
Incontinentia pigmenti, *557*
Independent assortment
  vs. linkage, 135–136
  Mendel's law of, 48–53, *49–50, 51,* 54–55, *55,* 130
Inducer molecules, 356, *357*
Inducible gene, 356, *357*
Inducible operon, 370–371
Induction
  cellular, 587
  lac operon, 358–359, *360, 361*
Influenza virus, *207, 208*
  H1N1 strain of, 213–214, *214*
Ingroup, 681–682
Inheritance, 10–11
  autosomal dominant, 554–556, *555*
  autosomal recessive, 553–554, *554*
  blending hypothesis of, 40
  chromosome theory of, *53,* 53–56, *54, 55*
  extranuclear
    chloroplast, 110–114, *111, 112, 113, 114*
    mitochondrial, *114,* 114–116, *115, 116*
  genetic disease, *553*
  of genetic disease, 552–558
  maternal, *123,* 123–125, *124*
    chloroplast, 111–112, *112, 113*
    mitochondrial, *115,* 115–116
  Mendelian (*See* Mendelian inheritance)
  particulate theory of, 46
  patterns of, *56,* 56–57
  polygenic, 651–656, *652, 654*
  pseudoautosomal, 79
  X-linked, 79–83, *80–81, 82*
  X-linked dominant, 556, *557*
  X-linked recessive, 556, *556*
Inhibitor molecules, 356, *357*
Initiation stage
  in transcription, *301,* 301–302
  in translation, *344,* 344–346, *345*
Insects, sex determination in, 72, *72*
In situ hybridization, 501–502, *502*
Insulators, 394–396, *395*
Insulin, bacterial production of, 1, 475, *476*
Insulin-like growth factor, 482, *482*
Integrase, 209, *210,* 522, *522*
Intelligence quotient (IQ), 662, *662*
Intergenic mapping, 137–145, *138, 139, 141, 142, 144,* 218
Intergenic suppressor mutation, 415, *416*
Internal nuclear matrix, 258, *258*
Interphase, 24, *25, 28, 259,* 259–260
Intervening sequences (introns), 312, 314, *315*
Intragenic mapping, *218,* 218–222, *221*
Intragenic suppressor mutation, 415, *416*
Introns, 312
  removal of, 314, *315*
  self-splicing of, 314, *315*
Inuit, 618–619
Inversion, 158, *158, 162,* 162–164, *163*
Invertebrate development. *See* Developmental genetics, invertebrate; *Drosophila melanogaster*
Inverted repeats, 519, *520*
In vitro fertilization, 563
Ion exchange chromatography, A-3–A-4
Iron regulatory protein, 403, *404*
Iron response element, 403, *404*
Iron uptake, 403, *403, 404*
Isoelectric focusing, 537–538
Isogamous species, 34
Itakura, Keichi, 475
Ivanovski, Dmitri, 206

J

Jacob, François, 193, 303, 358, 360
Jacobs syndrome, *170*
Jeffreys, Alec, 637
Jellyfish, *GFP* gene of, *3,* 3–4
Johannsen, Wilhelm, 46

K

Karpechenko, Georgi, 177–178
Karyotype, 21, *22,* 156, *157*
Kimura, Motoo, 688
Kinetochore, 24, *26, 27,* 33, *33,* 251, *251*
Klebsiella, 543, *544*
Klinefelter syndrome, *170*
Knockin mouse, *481,* 481–482
Knockout mouse, 481–482, 535
Knudson, Alfred, 570
Kornberg, Roger, 255, 310
Kraatz, Ernst Gustav, 595
Krüppel gene, 591, *592*
Kuru, *564*

L

lacA gene, 358, 360
lacI gene, 358, *359*
  mutations in, 359, 360–362, *361, 362–363*
Lack, David, 624
lacO gene, 364
  mutations in, 365–366, 414
lac operator site, 462
lac operon, 358–367
  activator protein regulation of, 364–365, *365*
  CAP site of, 358, *359*
  glucose effect on, 364–365, *365*
  induction of, 358–359, *360, 361*
  lacI gene of, 359, 360–362, *361, 362–363, 364*
  lacO gene of, 364
  lactose effect on, 359–361, *361*
  operator site of, 358, *359,* 365–366, *366, 367*
  organization of, 358, *359*
  repressor protein regulation of, 358–359, *360,* 360–362, *361, 362–363, 364*
lac repressor, 358–359, *360, 365,* 365–366, *367*
β-Lactamase, 445
Lactoferrin, transgenic production of, 482, *482*
Lactose metabolism, 358, *359. See also lac* operon
lacY gene, 543, *544*
lacZ gene, *195,* 195–196, *196,* 447–449, *448*
Lagging strand, *275,* 276, 277–278, *278, 279*
Lamarck, Jean Baptiste, 418–419
Lander, Eric, 653
Lathyrus odoratus, 102–103, *103*
Law of independent assortment, 48–53, *49–50, 51,* 54–55, *55,* 130
Law of segregation, 44–48, *47, 49,* 54, *54*
Leading strand, *275,* 276, 277–278, *278, 279*
Leber hereditary optic neuropathy, 116, *116*
Leder, Philip, 337
Lederberg, Esther, 188, 420
Lederberg, Joshua, 187, 188, 196, 420
Left-right axis, *590,* 591
Leptotene, 30, *31*
Lesch-Nyhan syndrome, *90,* 581
Lesion-replicating DNA polymerase, 284, *284,* 437
Lethal allele, *89,* 100–101, *101*
Leucine zipper motif, 382, *382*
Lewis, Edward, 587, 596
Light microscope, 21, *22*
Lineage diagram, 597–598, *598*
Linkage, 129–131
  Bateson and Punnett experiment and, 130, *130*
  chi square analysis for, 135–136
  crossing over and, 131, *131,* 136, *137*
  in dihybrid cross, 140–143, *141, 142*
  gene distance and, 132–133, *134, 139,* 139–140
  vs. independent assortment, 135–136
  in trihybrid cross, 143–144, *144*
  of X-linked genes, 131–132, *132, 133*
Linkage disequilibrium, 560
Linkage groups, 130
Linkage mapping, 138, *138,* 503–506, *510. See also* Genetic mapping
  microsatellites for, *505,* 505–506, *506*
  molecular markers for, *503,* 503–505, 653, *655,* 655–656
  of quantitative trait loci, 653, *655,* 655–656
Linker region, of DNA, 255–256, *355–356*
Lipids, 4
Lipman, David, 544
Liposomes, 492, *493*
Livestock, transgenic, 482, *483*
Local population, 615, *615*
Locus (loci), 23, *23,* 500
Lod score method, 505
Long interspersed elements, 522
Long terminal repeats (LTRs), 520, *520*
Loss-of-function allele, 14, 52, 103–105
Loss-of-function mutation, 14, 554
Loss of heterozygosity, 576–577, *577*
Low stringency condition, 460
Luria, Salvadore, 419
Lyell, Charles, 673
Lymnaea peregra, 123, 123–125, *124*
Lymphocytes, human immunodeficiency virus invasion of, 214, *214*
Lyon hypothesis, *76,* 76–77
Lysogenic cycle, 209, 212, *213*
Lysosome, 20, *20*
Lysozyme, 330, *332*
Lytic cycle, 212, *213,* 215, *216*

M

Mackay, Trudy, 657
MacLeod, Colin, 227
Macroevolution, 672. *See also* Evolution; Microevolution
Macromolecule, 4, *5*
Mad cow disease, *564*
Magnesium hypothesis, of DNA, 236
Malaria, 95–96, *518*
Malécot, Gustave, 631–632
Malthus, Thomas, 673
Manx cat, 100, *101*
Map distance, 140, 220
Mapping. *See* Chromosome mapping; Genetic mapping; Linkage mapping; Physical mapping
Mapping functions, 142, *142*
Map units (mu), 140
Marfan syndrome, *555,* 556, 581
Margulis, Lynn, 117
Mass spectrometry, *538,* 538–539
Maternal age, chromosome abnormalities and, 170–171, *171*
Maternal effect, *123,* 123–125, *124*
Maternal effect genes, *591,* 591–593, *592, 595*
Maternal inheritance
  in chloroplast, 111–112, *112, 113*
  in mitochondria, *115,* 115–116
Maternal myopathy and cardiomyopathy, *116*
Mather, Kenneth, 664
Matrix-attachment region, 258, *258*
Matthaei, J. Heinrich, 335

Maturases, 314
Maxam, Allan, 464
Maximum likelihood method, 684
Mayer, Adolf, 206
Mayr, Ernst, 675, 676
McCarty, Maclyn, 227
McClintock, Barbara, 136
McClung, Clarence, 72
McKusick, Victor, 630
Meadowlark, *674,* 674–675
Mean, 648–649
Mean fitness of population, 621–622
Mediator, 310, 383–384, *384*
Medicine, biotechnological production of, 475, *476,* 482, *482, 483, 575*
Megasporocyte, 36, *36*
Meiosis, 30–33
  anaphase of, *32,* 33
  crossing over during (*See* Crossing over)
  metaphase of, 31, *32,* 33
  vs. mitosis, 33, *34*
  prometaphase of, 31, *32*
  prophase of, 30–31, *31, 32*
  telophase of, *32,* 33
Mello, Craig, 400
Mendel, Gregor, 10, 13, 41, *41*
  experiments of
    characters studied in, 43, *44*
    pea plants for, 41–43, *42, 43,* 102
    single-character, 44–46, *45*
    two-character, 48–51, *49–50*
Mendelian inheritance
  ABO blood group, 96–98, *97*
  dominant/recessive, *90,* 90–92, *91*
  extensions of, 88–105
  gene interactions and, *102,* 102–105, *103, 104*
  incomplete dominance, *94,* 94–95, *95*
  law of independent assortment in, 48–53, *49–50, 51,* 54–55, *55,* 130
  law of segregation in, 44–48, *47, 49,* 54, *54*
  overdominance, *95,* 95–96, *96*
  pea plant experiments in, 41–44, *42–44*
  sex-influenced, 98–100, *99*
  sex-limited, 98–99, *100*
  simple, 88–89, *89*
Merozygote, 362
Meselson, Matthew, 270, 303
Messenger RNA. *See* RNA (ribonucleic acid), messenger (mRNA)
Metabolism, inborn errors of, 325
Metaphase
  chromosome compaction at, 260–261, *261, 262, 263*
  of meiosis, 31, *32,* 33
  of mitosis, *26, 28,* 29
Methylation, DNA, 239, 391–394, *392, 393, 394*
  in cancer, 574
  de novo, 392–393, *394*
  imprinting and, 119–120, *120, 121*
  maintenance, 392–393, *394*
Methyl-CpG-binding proteins, 392, *393*
Meyerowitz, Elliot, 604
micF gene, 372, *372*
Mic2 gene, 79, *79*
Microevolution, 619–620, *620,* 672
  genetic drift in, *620,* 628–629, *629, 630*
  migration in, *620,* 630–631
  mutation in, *620,* 634–635 (*See also* Mutation[s])
  natural selection in, *620,* 620–628 (*See also* Natural selection)
  nonrandom mating in, *620,* 631–633, *632*
Microinjection, in transgenic plant production, 488
Microsatellites, *505,* 505–506, *506,* 636
Microscopy, A-1–A-2
Microtubule, 26–27, *27*
Microtubule-organizing centers, 26
Migration, *620,* 630–631
Miller, Oscar, Jr., 348
Minisatellite, 636
Minute unit, in conjugation studies, 195, *195*
Mirabilis jalapa, 94, *94,* 111, *112, 113*
Mismatch repair, *431,* 433–434, *435*
  gene conversion by, 291–293, *292*
Missense mutation, 412, *413,* 568, *568*
Mitochondria, 20, *20*
  circular chromosomes of, 114–115, *115*
  DNA of, 115, *115*
  extranuclear inheritance in, 114–116
    maternal inheritance and, *115,* 115–116
    mutation and, 116, *116*
  nucleoid of, 114, *114*
  origin of, 116–117, *117*
Mitochondrial myopathy, *116*
Mitosis, 25, *25,* 27–30, *28*
  anaphase of, *28,* 29
  crossing over in, 145–147, *146*
  vs. meiosis, 33, *34*
  metaphase of, *26, 28,* 29
  outcome of, 29
  prometaphase of, 27–29, *28*
  prophase of, 27, *28*
  telophase of, *28,* 29
Mitotic spindle, 26–27, *27, 32*
MN blood group, 618–619
Model organism, 12, *13*
Moderately repetitive sequence, 252–253, *253*
Molecular clock, 688–690, *689*
Molecular evolution, 672–673, 686–691
  ancestral gene in, 686–687, *687, 688*
  chromosome changes in, 690, *690*
  neutral changes in, 688
  synteny groups in, 691, *691*
  timescale of, 688–690, *689*
Molecular genetics, 14
Molecular markers, *503,* 503–505, 653, *655,* 655–656
Molecular pharming, 482, *482*
Molecular profiling, 575–576
Monoallelic expression, 118
Monod, Jacques, 358, 360
Monohybrid cross, 44–46, *45*
Monomorphic strain, 657
Monophyletic group, 680, *680*
Monosomy, *168, 169,* 169–170
Monozygotic twins, 552
Moore, Graham, 691
Morgan, Thomas Hunt, 79–80, 131–132
Morphogen, 586–587, *587*
Morphological trait, 7
Morphs, 8
Mosaicism, 175–176, *176,* 417
Mosquito
  DDT resistance in, 623, *623*
  GFP gene in, *3,* 4
Motif, 381–382, *382,* 543
Mouse, *13, 518*
  gene replacement in, 479–481, *480*
  GFP gene in, *3,* 3–4
  imprinting in, *118,* 118–119, *119*
  knockin, *481,* 481–482
  knockout, 481–482, 535
  transgenic, *477,* 477–483, *480, 481*
Mouse model, 481–482
M phase, of cell cycle, 24, *25*
mRNA. *See* RNA (ribonucleic acid), messenger (mRNA)
Mullis, Kary, 450
Multicellularity, 23
Multinomial expansion equation, 60
Multiplication method, 51
Muscular dystrophy
  Duchenne, 82, *82, 556*
  myotonic, 424, *425*
Mutagen, *427,* 427–430, *428, 429*
  testing for, *429,* 429–430
Mutagenesis, site-directed, 466–467, *467*
Mutation(s), 8–9, 14, 411–438
  beneficial, 414, 623, *623*
  coding sequence effects of, 412–414, *413*
  conditional, 414
  deamination and, *421,* 422, *422*
  definition of, 8–9, 411
  deleterious, 414
  depurination and, *421,* 421–422, *422*
  dominant-negative, 91, 555–556
  down promoter, 414
  dynamic, 424–425
  fluctuation test of, *419,* 419–420
  frameshift, 333–334, *334,* 412, *413*
  frequency of, 429
  gain-of-function, 91, 555, 566, *566,* 596, *596*
  genotype effects of, 414
  germ-line, 417–418, *418*
  homoallelic, 220
  hot spots for, 221, *221,* 424–425, *425, 426*
  human disease and, 424–425, *425*
  induced, 421, *421, 427,* 427–430, *428, 429*
  lacI gene, 359, 360–362, *361, 362–363, 364*
  lacO gene, 364, 365–366, 414
  lethal, 414
  loss-of-function, 14, 554, 596
  missense, 412, *413,* 568, *568*
  mitochondrial, 116, *116*
  Neurospora, 325–327, *326*
  neutral, 412, 414, 688
  nonneutral, 688
  nonsense, 412, *413*
  outside coding sequence, 414, *414*
  oxidative stress and, 424, *424*
  phage T4, 215–218, *216, 217,* 333–334, *334*
  phenotype effects of, 414–415
  point, 412, *413*
  polar, 412, 439
  in populations, 634–635
  random nature of, 418–421, *419, 420*
  rate of, 429
  repair of (*See* DNA repair)
  replica plating of, *420,* 420–421
  reversion, 414
  silent, 412, *413*
  somatic, 417–418, *418*
  spontaneous, *421,* 421–425
  suppressor, 414–415, *416*
  tautomeric shift and, *421,* 422–424, *423*
  transition, 412
  translesion DNA polymerase replication of, 437
  transposable elements and, 522, 524–525, *525*
  transversion, 412
  up promoter, 414
Mutation frequency, 429
Mutation rate, 429, 634
Mycobacterium tuberculosis, 518
Mycoplasma genitalium, 518
Myoclonic epilepsy with ragged-red muscle fibers, *116*
MyoD gene, 601
Myogenic bHLH proteins, 601–602, *602*
Myotonic muscular dystrophy, 424, *425*

N

*Nae*I, *447*
Nägeli, Carl, 54, 226
Narrow-sense heritability, 659
Nathans, Werner, 446
Natural selection, 10, 673
  balancing selection in, *625,* 625–628
  Darwinian fitness and, 621
  directional selection in, 621–623, *622, 623*
  disruptive (diversifying) selection in, 624, *624, 625*
  in Galápagos finches, 626–628, *627*
  negative frequency-dependent selection in, 626, *626*
  stabilizing selection in, 623–624, *624*
Negative frequency-dependent selection, 626, *626*
Neurofibromatosis I, *555*
Neurospora, 325–327, *326, 518*
Neutral mutation, 412, 414, 688
Neutral theory of evolution, 688
Next-generation sequencing technologies, 514–517, *515*
Nick translation, 432
Nilsson-Ehle, Herman, 652–653
Nirenberg, Marshall, 335, 337

Nitrogen mustard, 427, *427*
Nitrous acid, 427, *427*
Noll, Markus, 255
Nonallelic homologous recombination, 159–160, *160*
Nonautonomous element, 520
Noncomplementation, 217, *217*
Nondisjunction
  complete, 175
  meiotic, 174–175, *175*
  mitotic, 174, 175–176, *176*
Nonessential gene, 100
Nonhomologous end-joining protein, 606
Nonhomologous end joining repair, *431,* 434–436, *436*
Nonhomologous genetic recombination, 199
Non–long terminal repeats (non-LTRs), 520, *520*
Non-Mendelian inheritance
  epigenetic, *118,* 118–122, *119, 120, 122*
  extranuclear, 110–114, *113, 114*
  mitochondrial, 114–116, *115, 116*
Nonneutral mutation, 688
Nonparental cell, 131
Nonparental generation, 50
Nonrandom mating, *620,* 631–633, *632*
Nonrecombinant cell, 131
Nonsense codon, 327
Nonsense mutation, 412, *413*
Normal distribution, 647, *648,* 649, *649*
Norm of reaction, 9, 93, *93*
Northern blotting, 460, *460*
Nowell, Peter, 569
Nuclear genes, 111
Nuclear genome, 246
Nuclear lamina, 258, *258*
Nuclear matrix, 257–258, *258*
Nucleic acids, 4, 231. *See also* DNA (deoxyribonucleic acid); RNA (ribonucleic acid)
Nucleoid, 20, *20*
  chloroplast, 111, *111*
  mitochondrial, 114, *114, 115*
Nucleolus, 312, 342
Nucleoside, 231, *233*
Nucleosome, *254,* 254–255
  in chromatin remodeling, *387,* 387–388, 390, *390*
  30-nm fiber formation of, 256–257, *257*
  structure of, *254,* 254–256, *255–256, 261*
Nucleosome-free region (NFR), *389,* 389–390
Nucleotide(s), 5, 231, *231*
  linkage of, *233,* 233–234
  structure of, *232,* 232–233, *233*
Nucleotide excision repair, *431,* 432–433, *433*
Nucleus, 4, *5,* 20, *20,* 253–260
Null hypothesis, 60, 135
NusA, 306, *307*
Nüsslein-Volhard, Christiane, 591, 593
Nyren, Pal, 515

O

OCA2 gene, 21
Okazaki fragments, *275,* 278, *278*
Olson, Maynard, 508
ompF gene, 372
Oncogenes, 566, *566, 568,* 568–569
One-gene/one-enzyme hypothesis, 325–327, *326*
On the Origin of Species by Means of Natural Selection (Darwin), 673
Oocyte, 10, 21, 34, *35*
  Drosophila, 589–591, *590*
  morphogenic gradients of, 586–587, *587*
  totipotent, 485, *486*
Oogenesis, 35, *35, 123,* 123–125, *124*
Open complex, 302
Operon, 358, 370–371. *See also lac* operon; *trp* operon
Orangutan, 682–683, *684,* 690, *690*
Organelles, *20,* 20–21
Origin of replication, 272–275, *273*
  bacterial, *274*
  eukaryotic, *283,* 283–284, *284*
  plasmid, 445
Origin of transfer, 189, *190*
Ortholog, 543–544, *544,* 599, 686–687, *687, 688*
Oryza sativa, 518
Osteoporosis, *555*
Outbreeding, 631
Outgroup, 681–682
Overdominance, *89, 95,* 95–96, *96,* 625–628
Ovum, 10, 21, 34, *35,* 485, *486*
Oxidative stress, 424, *424*

P

Pachytene, 30, *31*
PAC vector, 508–509
Pair-rule gene, 593–594, *595*
Palindromic sequence, 447
Pangenesis, 40
Pan troglodytes, 518
Paracentric inversion, 162, *162, 163,* 164
Paralog, 103–104, 160, *160,* 544, 686–687, *687, 688*
Parapatric speciation, *677,* 678
Pardee, Arthur, 360
Parental cell, 131
Parental (P) generation, 44, *45–46*
P1 artificial chromosomes, 508–509
Parvovirus, *207*
Pászty, Chris, 482
Patau, Klaus, 77
Patau syndrome, *170*
Paternal leakage, 115–116
Paterson, Andrew, 656
Pattern, 586
  body, 586–587, 589–591, *590, 693,* 693–694
Pattern baldness, 99, *99*
Pauling, Linus, 234, *234,* 680
Pax6 gene, 694–695, *695*
Pea plants
  Mendel's study of, 13, 41–44, *42–44,* 102
  morphological features of, 42, *42*
Pearson, Karl, 648, 651
Pedigree analysis, *56,* 56–57, 553–556
  in Duchenne muscular dystrophy, 82, *82*
  in Ehler-Danlos syndrome, 581
  in familial hypercholesterolemia, 579–580
  in hemophilia A, 556, *557*
  in human pattern baldness, 99, *99*
  in Huntington disease, 554–556, *555*
  in Hurler syndrome, 581
  of inbreeding, 631–633, *632*
  in Lesch-Nyhan syndrome, 581
  linkage mapping for, 506, *506*
  in Sandhoff syndrome, 582
  in Tay-Sachs disease, 553–554, *554*
P element, 522, *524,* 524–525
Peptide bond, 329
Peptidyl (P) site, 343, *343*
Peptidyl transferase, 347, *347*
Peptidyl transfer reaction, 347
Pericentric inversion, 162, *162, 163,* 164
Personalized medicine, 576
p21 gene, 571
p53 gene, 571, *572*
P (parental) generation, 44, *45–46*
Phage. *See also* Bacteriophage
Phage fd, *207*
Phage λ, *207*
  reproductive cycle of, 209–212, *210–211*
  self-assembly of, 212
Phage Qβ, *207*
Phage T2
  DNA of, *229–230,* 229–231
  Escherichia coli infection with, *228,* 228–231, *229–230*
  reproductive cycle of, *228,* 229
  structure of, *228,* 228–229
Phage T4, 206, *207, 208*
  Escherichia coli infection with, 215–218, *216, 217*
  intragenic mapping in, *218,* 218–222, *221*
  mutations in, 215–218, *216, 217,* 333–334, *334*
  structure of, 215, *215*
Phenogram, 681, 682–683, *683*
Phenotype, 46
  aneuploidy and, *170,* 170–171
  balancing selection of, 625–628, *626, 627*
  continuous variation in, 647, *648, 652,* 652–653
  copy number variation and, 162
  Darwinian fitness and, 621
  directional selection of, 621–623, *622, 623*
  disruptive selection of, 624, *624*
  environment and, 657–658, *658,* 662
  epistasis and, 102–103, *103*
  gene duplication and, 160
  genetic variation and, 656–657, 658–661, *660–661*
  imprinting and, *118,* 118–119, *119*
  maternal effect and, *123,* 123–125, *124*
  mutation effects and, 414–415
  polygenic trait and, 653, *654*
  quantitative study of (*See* Quantitative genetics)
  stabilizing selection of, 623–624, *624*
  variegated, 112, *113*
Phenylalanine, 325, *325*
Phenylalanine hydroxylase, *90,* 93
Phenylketonuria (PKU), 9–10, *10, 90,* 92–93, *93,* 325, *325, 554*
Phosphodiester linkage, *233,* 233–234, 282
Photolyase, 430, *431*
Photoreactivation, 430, *431*
Phragmoplast, *29*
Phylogenetic tree, 679–685, *686*
  Bayesian method for, 684
  cladistic approach to, *681,* 681–682, *683*
  construction of, 681–683, *682*
  interpretation of, 680, *680,* 684
  maximum likelihood method for, 684
  phenetic approach to, 681, 682–683, *683*
Phylogeny, 679
Physical mapping, 507–512, *510*
  BAC vector for, 508–509, *509*
  chromosome walking for, 510–511, *511*
  contig construction for, *507,* 507–508, *508*
  PAC vector for, 508–509
  positional cloning for, 510–511, *511*
  YAC vector for, 508–509, *509*
PIBIDS, 433
Pigmentation, gene for, 7–8, *8*
PKU, 9–10, *10, 90,* 92–93, *93,* 325, *325, 554*
Plant(s)
  chromosome number variation in, *173,* 173–174, *174*
  colchicine-induced polyploidy in, 179, *179*
  development of, 602–605
    ABC model of, 604, *605*
    homeotic genes in, 604–605, *605*
    steps in, *603,* 603–604, *604*
  dioecious, 74, *74*
  euploidy variations in, *173,* 173–174, *174*
  genetically modified, 177–179, *178,* 488–491, *489, 490, 491*
  genetic variation in, 9, *9*
  height of, 657–658, *658*
  reproduction of, 35–37, *36, 42,* 42–43
  selective breeding of, 663, *664*
  sex determination in, 74, *74*
  sterility of, 174, 176
Plaque, bacteriophage formation of, 215–218, *216, 217*
Plasmids, 189–190, 445
  Agrobacterium, 476, 488, *489*
  origin of replication of, 445
Plasmodium falciparum, 95–96, *518*
Pleiotropy, 101–102
Plethodon, 252, *252*
Pneumocystis jiroveci infection, 214
Point mutation, 412, *413*
Polar mutation, 412, 439
Pollen grains, 36, 42, *42*
PolyA-binding protein, 399
Polyadenylation, 317, *318*
PolyA tail, 317, *318*
Polycistronic RNA, 217, 358
Polycomb genes, 597
Polydactyly, 91–92, *92*
Poly-dT primer, *449,* 449–450

Polygenic inheritance, 651–656
  environmental factors and, 652–653, *653, 654*
  heritability and, 656–663, *658, 660–661, 662*
Polymerase chain reaction (PCR), 450–455
  microsatellite amplification by, *505,* 505–506, *506*
  real-time, 453–455, *454*
  reverse transcriptase, 453, *453*
  technique of, 450–453, *451, 452*
Polymerization, of pollutants, 476–477
Polymorphism, 505–506, *506,* 615–616, *616*
Polynucleotide, *233,* 233–234
Polypeptide, 5. *See also* Protein(s)
  α helix of, 330–331, *333*
  amino acid sequences of, 327–332, *328, 329, 330, 331*
  amino-terminal end (N-terminus) of, 329
  β sheet of, 330–331, *333*
  carboxyl-terminal end (C-terminus) of, 329
  definition of, 327
  directionality in, 329, *330*
  folding of, 330–331, *333*
  primary structure of, 330, *332, 333*
  pyrrolysine incorporation into, 329
  quaternary structure of, 332, *333*
  regulation of (*See* Gene regulation)
  secondary structure of, 330–331, *333*
  selenocysteine incorporation into, 329
  synthesis of (*See* Translation)
  tertiary structure of, 331–332, *333*
Polyploidy, *168,* 169
  colchicine-induced, 179, *179*
  plant, *173,* 173–174
Polyribosome, 349
Polysome, 349
Polytene chromosomes, 171–173, *172*
Population, 615
  conglomerate, 630–631
  disequilibrium in, 619
  local, 615, *615*
  mean fitness of, 621–622
  size of, 615
Population genetics, 14, 614–637
  allele frequency in, 616–617
  bottleneck effect in, 629, *630*
  definition of, 614
  DNA fingerprinting in, 636–637, *637*
  exon shuffling in, 635, *635*
  founder effect in, 630, *630*
  gene flow in, 632
  gene pool in, 614–615
  genetic drift in, 628–630, *629,* 630
  genetic polymorphism in, 615–616, *616*
  genotype frequency in, 616–617
  Hardy-Weinberg equation in, 617–619, *618*
  heterozygote carriers in, 618–619
  horizontal gene transfer in, 636, *636*
  microevolution in, 619–620, *620*
  migration in, *620,* 630–631
  mutations in, *620, 633,* 633–635
  natural selection in (*See* Natural selection)
  nonrandom mating in, *620,* 631–633, *632*
  repetitive sequence changes in, 636
  variation sources in, *620, 633,* 633–637, *635, 636, 637*
Populus trichocarpa, 518
Positional cloning, 510–511, *511*
Positional information, 586–587
Position effect, 415, *417*
Positive interference, 145
Posttranslational covalent modification, 373, *536,* 537
Posttranslational gene regulation, 356, *356, 372,* 372–373
Postzygotic isolating mechanisms, 675, *675*
Prader-Willi syndrome, 120–121, *122*
Preimplantation genetic diagnosis, 563
Preinitiation complex, 310
Pre-mRNA, 314, *315*
  alternative splicing of, *396,* 396–398, *397, 398*
  capping of, 316–317, *317*
  polyA tail of, 317, *318*
  splicing of, 314–316, *316*
Prenatal testing, *562,* 562–563
Prezygotic isolating mechanisms, 675, *675*
Pribnow box, 303, *303*
Primary structure, 330, *332, 333*
Primase, 276, *276,* 278–279, *279*
Primates, 690, *690*
Primers
  in cDNA synthesis, *449,* 449–450
  in polymerase chain reaction, 450–453, *451*
  in site-directed mutagenesis, 466–467, *467*
Primosome, 278–279, *279*
Principle of parsimony, 682
Prions, 563–564, *564*
Probability, 57–62
  definition of, 58
Probe
  antibody, 460–461, *461*
  in colony hybridization experiment, *457,* 457–458
Processive enzyme, 282
Processivity, 282
Product rule, 58–59, 145
Proflavin, 427, *427*
Programmed cell death, 571, 586, *586*
Prokaryote, 20, *20*
Prometaphase
  of meiosis, 31, *32*
  of mitosis, 27–29, *28*
Promoter, *300,* 301
  in bacterial transcription, *303,* 303–304
  in eukaryotic transcription, 308–309, *309*
  operon, 358
  regulatory element relation to, 383
Proofreading, in bacterial DNA replication, *282,* 282–283
Prophage, 209
Prophase
  of meiosis, 30–31, *31, 32*
  of mitosis, 27, *28*
Protandrous hermaphrodite, *73,* 73–74
Protease, 228
Protein(s). *See also* Polypeptide
  activator, 356, *357,* 364–365, *365,* 383, *383, 384*
  allosteric, 359
  in bacterial DNA replication, *275,* 275–277, *276, 277*
  cellular, 4–5, *5*
  checkpoint, 572–573, *573*
  chloroplast, 111
  coactivator, 383, *383*
  definition of, 327
  DnaA, 274, *274*
  DNA-binding, *462,* 462–463, *463*
  DNase I footprinting of, 462–463, *463*
  in eukaryotic transcription, *309,* 309–310, *310*
  functional activity of, 96, *96*
  function of, 7, *8*
  gel retardation assay of, 462, *462*
  heterozygote phenotype and, 91, *91,* 96, *96*
  histone, *254,* 254–255
  homodimer, 96, *96*
  in homologous recombination, 291, *291*
  in lactose metabolism, 358–359, *359*
  recessive alleles and, *90,* 90–91
  regulation of (*See* Gene regulation)
  regulatory, 356, *357*
  repressor, 356, *357,* 383, *383, 384*
  RNA-binding, 462
  SMC, 262, *262*
  structure of, 329–332, *332, 333*
    prediction of, 546–547
  subunit composition of, 96, *96*
  synthesis of (*See* Translation)
  Western blotting for, 460–461, *461*
Protein microarrays, 539, *539*
Proteome, 4–5, 531
  size of, 536–537
Proteomics, 500, 531, 535–540
  gel electrophoresis for, *537,* 537–538
  mass spectrometry for, *538,* 538–539
  protein microarrays for, 539, *539*
Proto-oncogenes, 566, *566, 568,* 568–569
Protoplasts, 488
Prototroph, 187
Provirus, 212
Proximodistal axis, *590,* 591
Pseudoautosomal genes, 78–79, *79*
*Pst*I, *447*
Pugh, J. E., 392
Punnett, Reginald, 47, 102, 130
Punnett square, 47–48, 51–52, *52,* 99, *99*
Purines, 232, *232*
P value, 61
Pyrimidines, 232, *232*
Pyrococcus abyssi, 518
Pyrosequencing, 515, *516*
Pyrrolysine, 329, *331*

Q

Quantitative genetics, 646–666
  continuous phenotypic variation and, 647, *648*
  environment in, 656–658, *658*
  frequency distribution in, 647, *648,* 648–649, *649*
  heritability in, 656–663, *658, 660–661*
  overview of, 646–647, *647*
  polygenic inheritance in, 651–656, *652, 654*
  QTL mapping in, 653, 655–656, *656*
  regression analysis in, 651
  selective breeding and, *663,* 663–665, *664, 665*
  variable comparisons in, 649–651, *651*
Quantitative trait locus (QTL), 652
  mapping of, 653, *655,* 655–656
Quaternary structure, 332, *333*
Quinolones, 249–250

R

Radial loop domain, 257–258, *258, 261*
Radioisotopes, detection of, A-5–A-6
RAG1, 606, *607*
RAG2, 606, *607*
Random mutation theory, 420
Random sampling error, 58
Ras protein, 566, *568*
rb gene, 570–571, *571,* 572
Rb protein, 571, *571*
Reading frame, 328–329
Realized heritability, 665–666
RecA protein, 291, *291*
RecBCD protein, 291, *291*
Receptor
  glucocorticoid, 385–386, *386*
  transferrin, 403, *403, 404*
Recessive allele, *90,* 90–91
Recessive inheritance, *90,* 90–92, *91*
  autosomal, 553–554, *554*
  X-linked, 556, *556*
RecG protein, 291, *291*
Reciprocal cross, *82,* 82–83
  maternal effect genes and, 123
Reciprocal translocation, *164,* 164–165, *165, 167,* 167–168
Recombinant cell, 131
Recombination. *See also* Crossing over
  genetic, 51, 131, 287
  homologous, 199, 268, 287–293, *289, 290, 291*
    double-strand break model of, *290,* 290–291
    double-strand break repair by, 434–436, *435*
    gene conversion and, 291–293, *292*
    Holliday model of, 288–290, *289*
    nonallelic, 159–160, *160*
    proteins in, 291, *291*
    resolution in, 290
  illegitimate, 199
  mitotic, 145–147, *146*
  nonhomologous, 199
Recombination signal sequence, 606, *607*
Recombinant DNA, 443. *See also* Biotechnology; Gene cloning; Polymerase chain reaction (PCR)
5-α-Reductase, 98
Regression analysis, 651
Regulatory (control) elements, 380, 382–384, *383, 384*

- Regulatory sequences/elements, *300,* 301
- Regulatory transcription factors. *See* Transcription factor(s), regulatory
- Relaxase, 189, *189, 190*
- Relaxosome, 189, *190*
- Release factors, 348, *348*
- Repetitive sequence changes, 636
- Replica plating, *420,* 420–421
- Replication. *See* DNA replication
- Replication forks, 272–277, *273, 275*
- Replicative transposon, 520, *520*
- Replisome, 278–279, *279*
- Repressible enzyme, 370–371
- Repressible gene, 356, *357*
- Repressible operon, 370–371
- Repressor, 356, *357,* 380, *381,* 383, *383, 384*
  - splicing, 397–398, *398*
  - translation inhibition by, 371
- Reproduction
  - asexual, 23–24, *24*
  - Darwinian fitness and, 621
  - definition of, 19
  - flowering plant, 35–37, *36, 42,* 42–43
  - sexual, 34–36, *35*
- Reproductive cloning
  - mammalian, 483–485, *484, 485*
  - plant, 488–491, *489, 490*
- Reproductive isolation, 675, *675*
- Repulsion (*trans* configuration), 140
- Restriction fragment length polymorphisms, 503–505, *504*
- Restriction point, 24, *25*
- Retinoblastoma, 570–571, *571*
- Retroelement, 253
- Retrotransposition, 519, *519, 520,* 521–522, *522, 523*
- Retrotransposons, 519, *519,* 521–522
  - LTR, 520, *520*
- Retrovirus, 212
- Rett syndrome, *557*
- Reverse genetics, 599–600
- Reverse transcriptase, *210,* 212
  - for retrotransposition, 521–522, *522*
- R factors, 445
- RFLP (restriction fragment length polymorphism), 503–505, *504*
- *Rhagoletis pomonella,* 678
- Rhesus monkey, 682–683, *684*
- Rho (ρ), 306, *306*
- Ribonucleic acid. *See* RNA (ribonucleic acid)
- Ribose, 232, *232, 233*
- Ribosomal RNA (rRNA), 252–253, *302,* 312, *314,* 342
- Ribosome, *20,* 341–343
  - aminoacyl (A) site of, 343, *343*
  - bacterial, 342, *342*
  - decoding function of, 346–347
  - eukaryotic, 342, *342*
  - exit (E) site of, 343, *343*
  - peptidyl (P) site of, 343, *343*
  - structure of, 343, *343*
  - subunits of, 343, *343*
- Ribosome-binding site, *300,* 301
- Riboswitches, *373,* 373–374, *374*
- Ribozyme, 313
- Rice, 691, *691*
- Rich, Alexander, 240
- Richmond, Timothy, 254, 257
- Riggs, Arthur, 392, 475
- RNA (ribonucleic acid), 6, *6*
  - antisense, 372, *372*
    - in mRNA silencing, 400–401, *400–401*
  - double-stranded, mRNA silencing by, *399,* 399–401, *400–401*
  - folding of, 241
  - functions of, 302, *302*
  - messenger (mRNA), 6, *6,* 299, *302,* 324
    - AU-rich element of, 399, *399*
    - double-stranded RNA silencing of, *399,* 399–400, *400–401*
    - for gene cloning, *449,* 449–450
    - pre-, 314–317, *315, 316, 317, 318, 396,* 396–398, *397, 398*
    - real-time PCR for, 453–455, *455*
    - Shine-Dalgarno sequence of, 345, *345*
    - stability of, *373,* 398–399, *399*
    - translation of (*See* Translation)
    - 3′-untranslated region of, 399, *399*
    - micro (miRNA), *302,* 401–402, *402*
  - modification of (*See* RNA modification)
  - Northern blotting for, 460, *460*
  - nucleotide structure of, *232,* 232–233, *233*
  - polycistronic, 217, 358
  - reverse transcriptase PCR for, 453, *453*
  - ribosomal (rRNA), 252–253, *302,* 342
    - processing of, 312, *314*
  - short, *337,* 337–338, *338*
  - short-interfering (siRNA), 401–402, *402*
  - small cytoplasmic (scRNA), *302*
  - small nuclear (snRNA), *302*
  - small nucleolar (snoRNA), *302*
  - strand of, 231, *231*
  - structure of, 231, *231, 240,* 240–241, *241*
    - prediction of, *546,* 546–547
  - synthesis of (*See* Transcription)
  - synthetic, 334–338, *337, 338*
  - telomerase, *302*
  - transfer (tRNA), 240–241, *241, 302,* 338–341
    - aminoacyl, 340, *340*
    - anticodons in, 327, *328*
    - charged, 340, *340*
    - folding of, 240–241, *241,* 339
    - function of, 338–339, *339*
    - initiator, 344, *344*
    - isoacceptor, 341
    - processing of, 312–313, *314*
    - structure of, 339, *339*
  - viral, 231, *302*
- RNA-binding proteins, 462
- RNA copolymers, *337,* 337–338
- RNA editing, *313,* 317–318, *318,* 537
- RNA-induced silencing complex, 402, *402*
- RNA interference (RNAi), *399,* 399–402, *400–401, 402*
- RNA modification, 312–318
  - alternative splicing in, *396,* 396–398, *397, 398*
  - base modification in, *313, 318*
  - 5′ capping in, *313,* 316–317, *317*
  - cleavage processing in, 312–313, *313, 314*
  - 3′ polyA tailing in, *313,* 317, *318*
  - RNA editing in, *313,* 317–318, *318*
  - splicing in, *313,* 314–316, *315, 316*
- RNA polymerase, *301,* 301–302
  - in bacterial transcription, *304,* 304–305, *308*
  - in eukaryotic transcription, 307–308, *308*
  - sigma factor subunit of, 304, *304*
- RNA polymerase holoenzyme, *304,* 304–305
- RNase, 228
- RNaseP, *302,* 313, *314*
- RNA splicing, 312, *313,* 314–316, *315*
  - alternative, *396,* 396–398, *397, 398*
- Robertsonian translocation, 165, *165*
- Ronaghi, Mostafa, 515
- Rooster, 99, *100*
- Root meristem, 603, *603*
- Rous sarcoma virus, 569–570
- rRNA, 252–253, *302,* 312, *314,* 342
- Rubin, Edward, 482
- Russell, Liane, 76
- RuvABC protein, 291, *291*

S

- *Saccharomyces cerevisiae, 13, 518*
  - centromere of, 251
- RNA polymerase II of, *308*
- RNA splicing in, 396–397
- Sachs, Bernard, 553
- *Sac*I, *447*
- Sager, Ruth, 113
- Salamander, 252, *252*
- Samecnik, Paul, 338
- Sandhoff disease, *90,* 582
- Scaffold protein, 260, *261*
- Schmitz, Albert, 462
- Schramm, Gerhard, 231
- *Sciara coprophila,* 119
- Science, 15
- Scientific method, 15
- Scintillation counting, A-6
- Scrapie, *564*
- SDS-PAGE (polyacrylamide gel electrophoresis), 460
- Secondary structure, 330–331, *333*
- Sedimentation coefficient, A-3
- Segmentation genes, *593,* 593–594, *594, 595*
- Segment-polarity genes, 593–594, *595*
- Segregation, Mendel's law of, 44–48, *47, 49,* 54, *54*
- Selectable marker, 445
- Selection. *See* Natural selection
- Selection coefficient, 625
- Selection limit, 664–665
- Selective breeding, *663,* 663–666, *664, 665*
- Selenocysteine, 329, *331*
- Self-fertilization, 43
- Selfish DNA hypothesis, 522
- Self-splicing, 314, *315*
- Semilethal allele, 100
- Semisterility, 167
- Sense codon, 327
- Separation methods, *A–2,* A–2–A–4
- *SEP* genes, 574, *575*
- Sequence complexity, 252–253, *253*
- Sequence element, 543
- Sequence ladder, 465–466
- Sequence-tagged sites, 505–506
- Sequencing by synthesis, 515, 517
- Sequencing ladder, 466
- Severe acute respiratory syndrome (SARS), 214
- Severe combined immunodeficiency disease (SCID), 492, 493–495, *494*
- Sex chromosomes, 71–72, *72. See also* X chromosome(s); Y chromosome(s)
  - dosage compensation and, *75,* 75–78, *76*
- Sex determination, 71–75
  - environment and, *73,* 73–74
  - haplodiploid system and, 72, *73*
  - in plants, 74, *74*
- Sex-influenced inheritance, *89,* 98–100, *99*
- Sex-limited inheritance, *89,* 99–100, *100*
- Sex-linked genes, 78–79, *79. See also* X-linked genes; Y-linked genes
- Sex pili, 188, *189*
- Sexual dimorphism, 99, *100*
- Sexual isolation, *675*
- Sexual reproduction, 34–36, *35*
- Shared derived character, 681
- Sheep, cloning of, 3, *3,* 483–485, *484*
- Shepherd's purse, 104, *104*
- Sheppard, Philip, 624
- Shimomura, Osamu, 4
- Shine-Dalgarno sequence, 345, *345,* 374, *374*
- Shingles, 213
- Shoot meristem, *603,* 603–604, *604*
- Short interspersed elements, 522
- Short tandem repeats, *505,* 505–506, *506,* 636
- Shotgun sequencing, 512–514
  - double-barrel, 513–514
  - hierarchical, 512–513, *513*
  - whole-genome, 513, *513*
- Shull, George, 104
- Sickle cell disease, *95,* 95–96, *96, 554,* 625, *626*
  - missense mutation in, *413,* 414
  - mouse model of, 481–482
- Sigma (σ) factor, 304, *304*
- Silencer, 309, 380, *381*
  - transcription factor regulation of, 382–383, *383*
- *Silene latifolia,* 74, *74*
- Silent mutation, 412, *413*
- Similarity, gene, 544
- Simpson, George Gaylord, 676
- Single-nucleotide polymorphism (SNP), 615
- Single-strand binding protein, 276, *276, 291*
- Sister chromatid exchange, 287, *288*
- Sister chromatids, 24, *26, 27, 32,* 33
- Site-directed mutagenesis, 466–467, *467*
- Site-specific recombination, *606,* 606–607, *607*
- Small nuclear riboproteins (snRNPs), 314–316, *316*
- SMC proteins, 262, *262*

- Smith, Hamilton, 446, 499
- Smith, Michael, 466
- Snail
  - shell coiling of, *123,* 123–125, *124*
  - shell color of, 624, *625*
- *SNRPN* gene, 121
- Somatic cells, 10
  - X chromosome in, 77
- Somatic hypermutation, 607
- Sorghum, 691, *691*
- SOS response, 437
- Southern, Edwin, 458
- Southern blotting, 458–460, *459*
- Speciation, 676, *677*
  - allopatric, 676–678, *677*
  - parapatric, *677,* 678
  - sympatric, *677,* 678, *679*
- Species, 7. *See also* Population genetics
  - definition of, 673–674
  - developmental differences among, *692,* 692–695, *693, 695*
  - ecological factors and, 675–676
  - evolutionary relationships among, 676, 679–685, *680, 682, 683, 686*
  - heterogamous, 34, 114
  - isogamous, 34
  - molecular features of, 675
  - morphological traits of, *674,* 674–675
  - morphs of, 8
  - origin of, 673–679, *674, 675, 677, 679*
  - reproductive isolation of, 675, *675*
- Species concept, 676
- Spectroscopy, A-5
- Spermatogenesis, 34–35, *35*
- Sperm cells, 10, 21, 34, *35*
- S phase, of cell cycle, 24, *25*
- Spinal and bulbar muscular atrophy, 424, *425*
- Spindle pole, 26, *27*
- Spinocerebellar ataxia, 424, *425*
- Spliceosome, 314–316, *315, 316*
  - splicing factors effects on, 397–398, *398*
- Splicing factors, 397–398, *398*
- Spontaneous mutation, *421,* 421–425
- Sporophyte, 35–36, *36*
- *src* gene, 569–570
- SR proteins, 397
- Stabilizing selection, 623–624, *624*
- Stahl, Franklin, 270
- Standard deviation, 649, *649*
- Start codon, *300,* 301, 327, 328, *328*
- Statistics, 57–62, 648–651
  - binomial expansion equation in, 59–60
  - chi square test in, 60–62, *62*
  - frequency distribution in, 648–649, *649*
  - multinomial expansion equation in, 60
  - normal distribution in, 647, *648,* 649, *649*
  - probability in, 58
  - product rule in, 58–59
  - regression analysis in, 651
  - sum rule in, 58
  - variables in, 649–651, *651*
- Stem cells, 485, *485*
  - in disease treatment, *486,* 486–487
  - embryonic, 485, *486*
  - multipotent, 485, *486*
  - pluripotent, 485, *486*
  - unipotent, 485, *486*
- Stem-loops, 369, *369, 370,* 374, *374*
- Sterility, plant, 174, 176
- Stern, Curt, 146–147
- Steroid hormones, transcription factor binding of, 385–386, *386*
- Stigma (plant), *42,* 42–43
- Stop codon, *300,* 301, 327, *328*
- Strahl, Brian, 389
- Strasburger, Eduard, 54
- *Streptococcus pneumoniae, 226,* 226–228, *227*
- *Sturnella magna, 674,* 674–675
- *Sturnella neglecta, 674,* 674–675
- Sturtevant, Alfred, 123, 140–143
- Subcloning, 511
- Subspecies, 674
- Sugarcane, 691, *691*
- *Sulfolabus solfataricus, 518*
- Sum rule, 58
- Supercoiling, DNA, *248,* 248–250, *249*
- Supergroups, 684, *685*
- Supernatant, A-3
- Superoxide dismutase, 688–690, *689*
- Suppressor mutation, 414–415, *416*
- Sutter, Nathan, 663
- Sutton, Walter, 54
- Sweet pea, 102–103, *103,* 130, *130*
- Sympatric speciation, *677,* 678
- Synapomorphy, 681
- Synapsis, 30, *31*
- Synaptonemal complex, 30–31, *31*
- Synteny, 130
- Synteny groups, 691, *691*

T

- Tandem array (repeats), 251, 253, 636
- TaqMan detector, 454, *454*
- *Taq* polymerase, 451, *452*
- Target-site primed reverse transcription, 522, *523*
- TATA box, 308, *309,* 383
- Tatum, Edward, 187, 325–326
- Tautomer, 422, *423*
- Tautomeric shift, 422–424, *423*
- Tay, Warren, 553
- Tay-Sachs disease, *90,* 553–554, *554,* 561
- T DNA (transferred DNA), 488, *489*
- Telomerase, 285–286, *286, 287, 302*
- Telomerase reverse transcriptase (TERT), 286
- Telomeres, 164, 251, *251,* 285–286, *286*
- Telophase
  - of meiosis, *32,* 33
  - of mitosis, *28,* 29
- Temperature, sex determination and, 73, *73*
- Temperature-sensitive lethal allele, 100
- Temporal isolation, *675*
- Termination (ter) sequences, 279–280, *280*
- Termination stage
  - in transcription, *301,* 301–302
  - in translation, 344, *344,* 348, *348*
- Terminator, *300,* 301, 358
- Termocycler, 452–453
- Tertiary structure, 331–332, *333*
- Testcross, 52, 81, *139,* 139–140
- Testosterone, in clownfish, 73–74
- Tetrad, 30
- Tetraploidy, *168,* 169, 171, *172*
  - plant, 173, *173*
- Thalassemia, 558
- *Theridion grallator,* 615, *616*
- Therman, Eva, 77
- *Thermoplasma volcanium, 518*
- Thiamin (vitamin B$_1$), 373, *374*
- Thiamin pyrophosphate (TPP), 373, *374*
- *thi* operon, 373, *374*
- Thirty (30)-nm fiber, 256–257, *257, 261*
- *thMDi* operon, 373–374, *374*
- Thoma, Fritz, 256
- Thomson, James, 486
- Thymine, 5, 232, *232,* 235, *235*
- Thymine dimer, *428,* 428–429, 430, *431*
- Tissue plasminogen activator, 482, *482*
- T lymphocytes, human immunodeficiency virus invasion of, 214, *214*
- Tobacco mosaic virus (TMV), 206, *207, 208*
  - self-assembly of, 212
- Tobacco plant, chloroplast DNA of, 111, *112*
- Tonegawa, Susumu, 383
- Topoisomerase I, 249, *250*
- Topoisomerase II, 248–249, *250, 275,* 276, *276*
- Topoisomers, 248, *249*
- Totipotent cells, 485, *486,* 602
- Trait(s), 43, *44*
  - behavioral, 7
  - continuous, 647
  - definition of, 4, 7
  - discontinuous, 647
  - in *Drosophila,* 79–80, *80–81*
  - environment and, 9–10, *10*
  - expressivity of, 91–92
  - gene expression and, 7–8, *8,* 52
  - genetic variation and, 8–9, *9*
  - incomplete penetrance of, 91–92, *92*
  - meristic, 647
  - morphological, 7
  - organism level of, 7, 8, *8*
  - physiological, 7
  - polygenic, 651–656, *652, 654*
  - population level of, 7, 8, *8*
  - quantitative (complex), 646–647, *647* (*See also* Quantitative genetics)
  - threshold, 647
- *Trans*-acting factors, 309, 364
- Transcription, 6, *6,* 299, *300,* 300–302. *See also* Gene regulation; RNA modification
  - bacterial, *300,* 302–307
    - activator protein in, 356, *357*
    - closed complex in, *304,* 304–305
    - coding strand in, 305, *305*
    - consensus sequence in, 303–304
    - corepressor molecules in, 356, *357*
    - elongation/synthesis stage of, 305, *305*
    - inducer molecules in, 356, *357*
    - inhibitor molecules in, 356, *357*
    - initiation of, *304,* 304–305
    - intrinsic termination of, 306, *307*
    - Pribnow box in, 303, *303*
    - promoter in, *303,* 303–304
    - regulation of, 356, *356, 357*
    - repressor protein in, 356, *357*
    - rho (ρ)-dependent termination of, 306, *306*
    - rho (ρ)-independent termination of, 306, *307*
    - riboswitch in, 373, *373, 374*
    - RNA polymerase holoenzyme in, *304,* 304–305
    - sequence elements in, *303,* 303–304
    - sigma factor in, 304, *304*
    - small effector molecules in, 356, *357,* 364–365
    - start site in, 303, *303*
    - template strand in, 305, *305*
    - termination of, *306,* 306–307, *307*
  - base sequences in, *300,* 300–301
  - eukaryotic, 307–312
    - activation of, *390,* 390–391
    - basal, 308
    - *cis*-acting elements in, 309
    - combinatorial control in, 380–381
    - core promoter in, 308–309, *309*
    - down regulation of, 383
    - elongation stage of, *390,* 390–391
    - enhancers in, 308–309, 380, 382–383, *390*
    - general transcription factors in, *309,* 309–310, *310*
    - histone code in, 388–391, *389, 390*
    - initiation of, *309,* 309–310, *310,* 390, *390*
    - mediator in, 310
    - nucleosome-free region in, *389,* 389–390, *390*
    - preinitiation complex in, 310
    - regulatory transcription factors in, *380,* 380–386, *381*
    - RNA polymerases in, 307–308, *308,* 310–311, *311*
    - silencers in, 309, 380, 382–383
    - TATA box in, 308
    - termination of, 310–312, *311*
    - *trans*-acting factors in, 309
    - up regulation of, 383
  - stage(s) of, *301,* 301–302
    - elongation/synthesis, *301,* 301–302
    - initiation, *301,* 301–302
    - termination, *301,* 301–302
  - translation coupling with, 348–349, *349*
  - Z-DNA in, 240
- Transcriptional start site, 303, *303*
- Transcription factor(s), 301
  - in cell differentiation, 600–602, *602*
  - definition of, 380
  - general, *309,* 309–310, *310,* 380
  - regulatory, *380,* 380–386, *381*
    - coactivator interaction with, 383, *383*

- Transcription factor(s), *continued*
  - covalent modifications of, 384, *385*
  - dimerization of, 382
  - domains of, 381, *382*
  - mediator interaction with, 383–384, *384*
  - modulation of, 384, *385*
  - motif of, 381–382, *382*
  - protein-protein interactions and, 384, *385*
  - regulatory element recognition by, 382–383, *383*
  - repressor interaction with, 383, *383*
  - small effector molecule interaction with, 384, *385*
  - steroid hormone binding to, 385–386, *386*
  - structure of, 381–382, *382*
- Transcription factor IID, *309,* 309–310, *310,* 383, *383*
- Transcription-repair coupling factor, 436–437, *437*
- Transduction, 186, *186*
- Transferred DNA (T DNA), 488, *489*
- Transferrin receptor, 403, *403, 404*
- Transfer RNA. *See* RNA (ribonucleic acid), transfer (tRNA)
- Transformation, 186, *186*
  - in bacteria, *226,* 226–228, *227*
  - in gene cloning, 447–448, *448*
- Transgene, 474
- Transgenic livestock, 482, *483*
- Transgenic organisms, 474, *477,* 477–478
  - gene addition, *478,* 478–479, *479*
  - gene knockin, *481,* 481–482
  - gene knockout, 481–482, *482*
  - gene replacement, 478–481, *480*
- Transgenic plants, 488–491, *489, 490*
- Translation, 5–6, *6,* 299, *300,* 324–350, *328*
  - anticodons in, 327, *328*
  - bacterial, *344,* 344–349, *345, 346, 347, 348, 349*
    - regulation of, 356, *356, 373,* 373–374, *374*
  - cell-free system of, 334–337, *335–336*
  - codons in, 327, *328*
  - directionality of, 329, *330*
  - elongation stage of, 344, *344, 346,* 346–347
  - eukaryotic, *344,* 344–349, *346, 349*
  - genetic code and, 327–328, *328, 329*
  - initiation stage of, *344,* 344–346, *345*
  - peptidyl transfer reaction in, 347
  - ribosome in, 341–343, *342, 343*
  - stages of, *344,* 344–349
  - termination stage of, 344, *344,* 348, *348*
  - transcription coupling with, 348–349, *349*
  - tRNA in, 338–341, *339, 340, 341*
- Translational regulatory protein, 371
- Translesion synthesis, 437
- Translocation, 158, *158, 164,* 164–165, *568,* 569, *569*
  - balanced, 165, *166*
  - gene expression and, 415, *417*
  - reciprocal, *164,* 164–165, *165, 167,* 167–168
  - Robertsonian, 165, *165*
  - unbalanced, 165
- Translocation cross, 165
- Transmission genetics, 13–14
- Transposable elements, 253, 517, *519,* 519–525, *524,* 636
  - direct repeats of, 519, *520*
  - DNA sequences of, 519–520, *520*
  - insertion sequence of, 519, *520*
  - inverted repeats of, 519, *520*
  - mutation and, 522, 524–525, *525*
- Transposase, 519, 520–521
- Transposition, 517, 519–525
  - mechanisms of, 519, *519*
  - replicative, 519, *519, 520*
  - retro-, 519, *519, 520,* 521–522, *522, 523*
  - simple, 519, *519, 520,* 520–521, *521*
- Transposon
  - composite, 520, *520*
  - excision and insertion of, 520–521, *521*
  - replicative, 520, *520*
- Transversion, 412
- Trihybrid cross, 130, 131–132, *132, 133*
  - crossing over in, 143–144, *144*
- Trinucleotide repeats, 424–425, *425, 426*
- Triplet
  - in genetic code, 333–334, *334*
  - synthetic, 337, *337*
- Triplet-binding assay, 337–338, *338*
- Triple X syndrome, *170*
- Triploidy, *168,* 168–169
  - plant, 173–174, *174*
- Trisomy, *168, 169,* 169–170
- Trisomy 13, *170*
- Trisomy 18, *170*
- Trisomy 21, 9, *9, 170*
  - familial, 165, *166*
  - maternal age and, 170–171, *171*
- Trisomy X, 170
- *Trithorax* genes, 597
- *Triticum aestivum,* 173, *173, 652,* 652–653, *654*
- tRNA. *See* RNA (ribonucleic acid), transfer (tRNA)
- α-Tropomyosin, 397, *397*
- *trp* operon, 367–371
  - attenuation regulation of, 368–370, *369*
  - repressor protein regulation of, 368, *368*
- Trp repressor protein, 368, *368*
- True-breeding line/strain, 43
- Tryptophan, 367–371, *368, 369, 370*
- Tsien, Roger, 4
- Tumor-suppressor genes, 570–571, *576,* 576–577
  - functions of, 571–572, *572*
  - silencing of, 574
- Turner, Bryan, 389
- Turner syndrome, *170*
- Tus, *276*
- Twins, 552
- Two-dimensional gel electrophoresis, *537,* 537–538
- Two-hit hypothesis, 570–571, *571*
- Tyrosinase, *90*
- Tyrosinosis, 325, *325*

U

- *UBE3A* gene, 121
- Ultraviolet light, *427,* 428–429
- Unbalanced translocation, 165
- Up regulation, 383
- Uracil, 232, *232*

V

- Van Valen, Leigh, 676
- Variables, statistical, 649–651, *651*
- Variance, 649, 656–658
- Variant, 43, *44,* 46
- Varicella-zoster virus, 213
- Vector(s)
  - BAC, 508–509, *509,* 512–513, *513*
  - expression, *445*
  - in gene cloning, 444–445, *445,* 447–449, *448*
  - PAC, 508–509
  - recombinant, 447–449, *448*
  - T-DNA, 488
  - YAC, 508–509, *509*
- Venter, Craig, 499
- Vertebrate development, 599–602
  - homeotic genes in, *599,* 599–600, *600, 601*
  - transcription factor genes in, 600–602, *602*
- Vieira, Christina, 657
- Viral envelope, 207–208, *208*
- Virus(es), 206–222. *See also* Bacteriophage; Phage
  - in cancer, *568,* 569–570, *570*
  - capsid of, 207, *208*
  - capsomers of, 207, *208*
  - definition of, 206, 207
  - emerging, 213–214
  - envelope of, 207, *208*
  - genome of, 209
  - host range of, 207
  - hosts of, 207, *207*
  - latency in, 212–213
  - nucleic acid of, 209
  - reproductive cycle of, 209–215, *210–211, 213, 214*
    - host cell attachment in, 209, *210*
    - host cell entry in, 209, *210*
    - integration in, 209, *210,* 212
    - lysogenic, 209
    - lytic, 212
    - viral assembly in, *211,* 212
    - viral release in, *211,* 212
    - viral self-assembly in, 212
    - viral synthesis in, *211,* 212
  - RNA, *207,* 231
  - structure of, *207,* 207–209, *208*
- Vitamin D-resistant rickets, *557*
- V(D)J recombination, 606–607
- Vogelstein, Bert, 574
- Volkin, Elliot, 302–303

W

- Wallace, Alfred Russel, 620, 673
- Watson, James, 235–236, *236,* 514
- Weakness, neurogenic, *116*
- Weinberg, Wilhelm, 617
- Weintraub, Harold, 601
- Weischaus, Eric, 591, 593
- Weismann, August, 54, 226
- Western blotting, 460–461, *461*
- Wexler, Nancy, 560
- Wheat, 9, *9,* 173, *173, 652,* 652–653, *654,* 691, *691*
- White campion, 74, *74*
- Wild mustard plant, selective breeding of, 663, *664*
- Wild-type alleles, 90, *90,* 414
- Wilmut, Ian, 3, 483
- Wilson, Henry, 587
- Wobble base, 328, 341, *341*
- Wobble rules, 340–341, *341*
- Woese, Carl, 684
- Wollman, Elie, 193
- Woodcock, Christopher, 257
- Wright, Sewall, 628
- *WUS* gene, 603–604, *604*

X

- X chromosome(s), 22, *22,* 72, *72*
  - crossing over of, 131–134, *132, 133, 134*
  - in *Drosophila melanogaster,* 175–176, *176*
  - genes of, 78–79, *79*
  - inactivation of, *75,* 75–78, *76, 78*
  - multiple, 170
  - in somatic cells, 77
- Xeroderma pigmentosum, 433, *434,* 574
- X inactivation, 75–77, *76, 78,* 119
- X-inactivation center (Xic), 77
- *Xist* gene, 77
- X-linked genes, 78–79, *79,* 81
  - dominant, 556, *557*
  - in *Drosophila melanogaster,* 131–134, *132, 133*
  - linkage of, 131–132, *132, 133*
  - recessive, 556, *556*
  - transmission of, 79–83, *80–81, 82, 89*
- X-ray, *427,* 428
- X-ray crystallography, 254, *254*
- X-ray diffraction, 234–235, *235*

Y

- YAC (yeast artificial chromosome) vector, 508–509, *509*
- Yamanaka, Shinya, 487
- Y chromosome(s), 22, *22,* 72, *72*
  - genes of, 78–79, *79*
- Y-linked genes, 78–79, *79*
- *Yucca,* 675

Z

- Z chromosome, 75
- Z DNA, *239,* 239–240
- Zickler, H., 288
- Zinc finger motif, 381–382, *382*
- Zinder, Norton, 196
- Zoller, Mark, 466
- Zoo blot, 458
- Zuckerkandl, Emile, 680
- Zygotene, 30, *31*
- Zygotic genes, 594